485
9.11.76

THE JOURNAL OF EXPERIMENTAL EDUCATION

A QUARTERLY

September 1962 - June 1963

VOLUME XXXI

DEMBAR EDUCATIONAL RESEARCH SERVICES, Inc.

madison 3, wisconsin

TABLE OF CONTENTS

VOLUME XXXI

FALL 1962

WINTER 1962

SPRING 1963

The Journal of Experimental Education

PUBLISHED QUARTERLY

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 1, September 1962)

MATHEMATICAL-VERBAL ABILITY DIFFERENTIALS AND SOMATIC EXPRESSIONS OF SITUATIONAL ANXIETY

ROBERT L. MILLIKEN and BERNARD SPILKA
Purdue University University of Denver

ONE OF the most common problems confronting the counselor in an educational setting concerns students who appear to demonstrate marked ability discrepancies in different areas of academic endeavor. Probably the most frequent form of such ability differentials is that between verbal and mathematical skills. The number of studies and discussions on such difficulties is, needless to say, exceedingly great. Among the various approaches utilized has been the attempt to relate such verbal-quantitative aptitude differences to personality factors. To date, however, the investigations conducted appear to have produced more contradiction and conflict than agreement and support (2, 11, 12, 15). One reason for this may be that the emphasis has been on the general characteristics of personality rather than with immediate situational variables. In this regard, Dreger and Aiken (6) recently pointed to the probability that situational anxiety relating to numerical materials rather than general anxiety may be of considerable significance.

The theory underlying this view has been stated by Lecky (8), Sullivan (16) and Wilson (21) among others. In brief, "early failure experiences in handling numerical materials... has provided a basis for the conditioned association of anxiety and mathematical symbols." (14, p. 60). Evidence for such conditioned associations should be found on all levels of manifest anxiety, such as feeling-tone expressions (6), and behavioral, and somatic expressions of anxiety manifestations. The present study represented an attempt to assess this last form of situational anxiety expression.

Specifically it was hypothesized that somatic expressions of anxiety would be found largely during the testing of abilities in the deficient skill area.

Procedure

Subjects. Forty-eight Ss, 24 males and 24 females ranging in age from 17 to 23 years, volunteered for this study. All were either college freshmen or prospective freshmen students at the University of Denver. On the basis of scores on the Scholastic Aptitude Test, hereafter referred to as SAT (4), three groups of Ss were formed, each with eight males and eight females. These were: 1) a high verbal-low mathematical ability group, hereafter referred to as the V group; 2) a high mathematical-low verbal skills group, the M group; and 3) an equivalent group in which both verbal and mathematical scores appeared relative equal, the V-M group. SAT score differentials for the V and M groups were 100 points or more, whereas the V-M group differentials were always less than 40 points. In order to prevent extreme deficiencies from confounding the results, the lowest verbal score employed was 370 and the lowest mathematical score was 310. A summary of these data may be found in Table I.

In addition, to remove the effects of generalized anxiety, all Ss who received scores of 20 or higher on the Taylor Manifest Anxiety Scale (17), and above eight on any of the Spilka-Siegel scales of Somatic, Feeling-Tone, and Behavioral Anxiety (10) were eliminated from the study. The final sample of 48 Ss represented those of an initial group of 86 possible Ss who met the above criteria.

Materials and Apparatus—Mathematical and and verbal abilities were tested by employing a modification of the American Council on Education Psychological Examination (18). Approximately one-fifth of the test questions were chosen from each section of the ACE, and these were presented individually to each subject and graded as to difficulty. In order to obtain the somatic measures of anxiety, a Keeler polygraph was employed and five measures were recorded on the polygram. These were breathing depth, breathing rate, blood pressure, heart rate, and psychogalvanic skin response. The basic data analyzed was derived from both the first 30 seconds of testing at the beginning of each subtest of the ACE (the

initial testing data), and the last 30 seconds of each subtest (the final testing data). Each 30-second period was subdivided into six five-second intervals. A mean was obtained for each of these intervals, and weighted grand means (here termed Mean means) were computed for each 30-second period. A variance of these measures was also computed for each 30-second period. These are termed the Variance means. This form of analysis was employed in order to assess changes in both the level and variability of physiological reactivity since both kinds of change are often indicative of emotion.

Analysis—In order to analyze the findings, a Mixed-Model Type III analysis of variance design (9) was employed. This permitted assessment of the differences relative to the groups, sexes, and tests, and all interactions among these effects. The initial and final testing data were analyzed separately. Twenty such analyses of variance, resulting in 140 significance tests were thus computed. Whenever significance was obtained, the procedure developed by Tukey (17) for comparing specific means was utilized.

Results

Of the 140 affects tested, 32 attained significance beyond the five percent level of confidence. This is considerably above what might have been expected on the basis of chance alone. These findings are summarized in Table II. (Because of the great complexity of these data, only summary tables are presented with this article. Any and all specific results may be procured on request from the writers.)

Breathing Depth and Rate— Turning first to consideration of the breathing depth and rate measures, nine significant effects were obtained. Some of these may be simply explained by natural sex differences in vital capacity and tidal air exchange. Emotional variation in breathing rate and depth might be manifested by such indicators as increased depth of breathing, and irregularity are noted for all groups when taking the mathematics tests on the initial testing, thus emotion seems to be a correlate of the mathematics testing irrespective of group effects. Possibly the increased emotion due to test difficulty obscured any further differences during the final testing.

Despite the fact that a significant test by group interaction was noted for final testing on breathing rate, the variation between groups, it is felt, is too small to be interpreted reliably.

In brief, the breathing measures analysis suggests that where emotion may be operative, the mathematics testing is most likely to demonstrate its presence. Increased breathing depth and irregularity of depth was noted here.

Blood Pressure— The blood pressure measurements yielded eight significant effects. Specifically, one might expect emotion to be correlated with an increase in both the level and variability of the blood pressure readings. Findings in support of this hypothesis were evidenced, as with breathing depth, during the mathematics testing. Blood pressure readings tended to be elevated when the subjects took the mathematics portions of the ACE. This was, however, found only in the final testing period, suggesting that blood pressure changes may be more lasting and slower to arouse than variations in breathing. Also as expected, the high varbal group demonstrated the highest blood pressure readings during the mathematics testing and the lowest during the verbal testing. No significance, however, can be claimed between the two tests for the high mathematics group.

Some support may be lent to the notion that men are more proficient in mathematics and less skillful in language than females, as significant sex by tests interactions were found for both means and variances. The group measures revealed that the males were possibly less anxious on the mathematical tests and more anxious than the females on the verbal tests.

Again, though not completely clear, one can easily infer evidence from the blood pressure data that demonstrates the association of situational anxiety with low mathematical ability.

Heart Rate--Wiffers and other researchers (18) feel that heart rate is a reliable indicator of emotion. It was specifically hypothesized here that increases in heart rate and irregularity of heart rate might be revealed during the testing of the deficient ability. It should first be noted that despite the fact that all the differences observed are quite small, four significant effects were obtained. With the limitation of small differences in mind, evidence is again obtained to suggest the association of increased heart rate with the mathematics testing, in general, during the final measurement period. This corresponds to the normal association of increased heart rate and blood pressure observed by Cannon and others. The significant tests by groups interactions do not reveal any clear or reliable pattern which might support or negate the foregoing hypothesis.

PGR— Eight significant effects were observed in the Psychogalvanic Response measures. In general, high PGR deflections on the polygram are associated with emotion. Irregularity of the pattern is also suggestive of anxiety. The results of the group by tests interaction supported the

hypothesis clearly. In both the High Verbal and High Mathematical groups, the testing of the deficient ability produced greater and more variable PGR deflections in both the initial and final testing sessions. Significant sex by group interactions were also obtained for both means and variances in the initial and final measurement periods. The group by sex differences are the same relationship as those observed for blood pressure, again suggesting the often hypothesized higher skills of males in mathematics, and females in verbal tasks.

Discussion and Conclusions

Although the findings reported here do provide support for the hypotheses, the results are by no means unequivocal. Sex differences and often times extremely small variation among the groups suggest a somewhat tentative acceptance of these results, especially pertaining to their practical significance, as opposed to their statistical significance.

One of the fundamental questions which may be asked of this research concerns the meaningfulness of investigating physiological correlates of emotional states. First, certain of these indices are highly controversial, such as PGR and Blood Pressure, despite the fact that the measurement of blood pressure by the cuff and ascultatory method appears to be the only procedure employed in the investigation of emotion and blood pressure. Wiggers (18) cites an impressive literature questioning the validity of the technique. The results obtained here, however, do seem to be generally supportive of the construct validity of this method. The PGR is also under attack as to its validity and reliability (1, 7, 12), yet it appears to be the most frequently employed somatic procedure for the assessment of emotion.

Second, there is no doubt of the threat-value of tests of a deficient ability (6), when evaluation is by verbal report. Phenomenological and physiological indicators of emotion, however, have not been satisfactorily correlated. Since the present study ignored the phenomenal aspect of the behavioral variable— namely ability discrepancies on performance tests— the broader questions of relating somatic, behavioral, and feeling-tone aspects of anxiety needs to be studied. In the present study, the ability discrepancies of the SAT appear to be verified, thus the research has in effect correlated behavioral and somatic variables. The feeling-tone, somatic, and behavioral relationships have not been assessed, and may well account for the discrepancies noted.

Further research is obviously needed to determine with more assurance if situational anxiety exists to a greater extent for non-preferred, low ability subject areas than for preferred, high ability subjects. The present study and that of Dreger and Aiken do, however, provide support for the meaningfulness of this quesion.

REFERENCES

1. Abel, T. "Attitudes and the Galvanic Skin Reflex," Journal of Experimental Psychology (XIII, 1930), pp. 47-50.
2. Altus, W. D. "Personality Correlates of Q-L Variability on the ACE," Journal of Consulting Psychology (XVI, 1952), pp. 284-91.
3. Cannon, W. B. Bodily Changes in Pain, Hunger, Fear, and Rage (Second Edition), New York, Appleton, 1929.
4. College Entrance Examination Board No. 2, Princeton, New Jersey, College Board Corp., 1955.
5. Darrow, C. W. "Multiple Physiological Changes as a Basis for Differentiating Emotional Responses," Psychological Bulletin (XXXIII, 1936), pp. 73-94.
6. Dreger, R. M. and Aiken, L. R. Jr., "The Identification of Number Anxiety in College Population," Journal of Educational Psychology (XLVIII, 1957) pp. 344-51.
7. Jones, H. "The Galvanic Skin Reflex as Related to Overt Emotional Expression" Journal of Psychology (XLVII, 1935), pp. 241-51.
8. Lecky, P., and Wilson, C. M., Self-Consistency: A Theory of Personality, New York, Island Press, 1945.
9. Lindquist, E. G., Design and Analysis of Experiments in Psychology and Education, Boston, Houghton-Mifflin, 1953.
10. Miller, Lee, Spilka, B., and Pratt, S., "Manifest Anxiety and Hostility in 'Criminally Insane' Patients," Journal of Clinical Experimental Psychopathy, (XXI, 1960) pp. 41-48.
11. Munroe, Ruth L. "Rorschach Findings on College Students Showing Different Constellations of Subscores on the ACE," Journal of Consulting Psychology (X, 1956), pp. 301-16.
12. Pemberton, C. L. "Personality Inventory Data Related to ACE Subscores," Journal of Consulting Psychology (XV, 1951), pp. 160-62.
13. Schlosberg, H. "Three Dimensions of Emotion," Psychological Review (LXI, 1954), pp. 81-88.
14. Spilka, B., "Numerical-Verbal Ability Differentials; a Theory and Research Program," Psychological Newsletter (X, 1958)

pp. 48-55.
15. Spilka, B. and Kimble, Gloria. "Personality Correlates of Q-L Differentials on the ACE," Journal of Consulting Psychology, (XX, 1958), p. 142.
16. Sullivan, H. S. Conceptions of Modern Psychiatry, Washington, D. C., William Alanson White Psychiatric Foundation, 1947.
17. Taylor, Janet A. "A Personality Scale of Manifest Anxiety," Journal of Abnormal Social Psychology (XLVIII, 1953), pp. 285-90.
18. Thurstone, L. L. and Thurstone, Thelma C. American Council on Education Psychological Examination for College Freshmen (1950 Ed.), Los Angeles, California, Educational Testing Service.
19. Walker, Helen M. and Lev, J. Statistical Inference, New York: Holt, 1953.
20. Wiggers, C. J. Physiology in Health and Disease (5th Ed.), Philadelphia: Lea and Febiger, 1949.
21. Wilson, C. M. Teaching the New Arithmetic, New York: McGraw-Hill, 1951.

TABLE I

MEANS AND STANDARD DEVIATIONS FOR AGES AND SAT SCORES BY GROUPS AND SEXES* VARIABLE

Group	Age		SAT M-Score		SAT M-Score	
	Mn	SD	Mn	SD	Mn	SD
High Math						
Male	19.1	1.62	633.0	88.0	438.9	53.1
Female	18.3	.66	540.8	59.1	412.8	45.6
Equivalent						
Male	19.2	1.71	494.8	52.9	491.0	55.5
Female	18.1	.60	473.8	22.8	472.3	13.2
High Verbal						
Male	18.5	.50	397.1	60.6	570.0	66.3
Female	18.4	1.50	425.0	84.8	591.9	66.8

* Eight subjects of each sex were included within each group, thus each group had 16 subjects.

TABLE II

SIGNIFICANCE OF ANALYSIS OF VARIANCE ON PHYSIOLOGICAL MEASURES

Variable	*Breathing Depth*		*Breathing Rate*		*Blood Pressure*		*Heart Rate*		*PCR*	
	In.	Fin.	In.	Fin.	In.	Fin.	In.	Fin.	In.	Fin.
				Mean Means						
Between Groups	--	--	--	--	.05	--	--	--	--	--
Between Sexes	.01	.05	--	--	--	--	--	--	--	--
G x S	--	--	--	--	--	--	--	--	.01	.01
M vs V	.05	--	--	--	--	.01	--	.05	--	--
(M - V) x Group	--	--	--	.05	--	.01	.05	--	.01	.05
(M - V) x Sex	.01	--	--	.05	.05	--	--	--	--	--
(M - V) Sex x Group	--	--	--	.05	--	--	--	--	--	--
				Variance Means						
Between Groups	--	--	--	--	.05	--	--	--	--	--
Between Sexes	.05	.05	--	--	--	--	--	--	--	--
G x S	--	--	--	--	--	--	--	--	.01	.05
M vs V	.05	--	--	--	--	.01	--	.05	--	--
(M - V) x Group	--	--	--	--	--	.05	.05	--	.01	.05
(M - V) x Sex	.01	--	--	--	.05	--	--	--	--	--
(M - V) x Sex x Group	--	--	--	--	--	--	--	--	--	--

TABLE III

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
BREATHING DEPTH-MEAN MEANS

	High Verbal Group			Equivalent Group			High Math Group		
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total
				Initial Testing					
Male									
Mean	19.1	17.2	36.3	15.3	13.4	28.7	15.2	14.5	29.7
S. D.	7.22	5.24	12.35	4.46	4.62	9.00	5.14	4.39	9.49
Female									
Mean	9.9	11.7	21.6	12.7	12.6	25.3	11.8	10.9	22.6
S. D.	5.36	5.35	10.38	3.58	3.66	7.33	3.40	2.82	5.95
Total									
Mean	14.5	14.4	29.0	14.0	13.0	27.0	13.5	12.7	26.2
S. D.	7.84	5.97	13.57	4.25	4.28	8.39	4.68	4.12	8.68
				Final Testing					
Male									
Mean	18.1	16.5	34.6	14.9	14.8	29.6	14.4	13.6	28.0
S. D.	5.30	5.13	10.26	4.37	3.82	8.14	5.65	4.48	10.01
Female									
Mean	10.0	11.0	21.0	12.2	11.4	23.6	13.2	12.7	25.9
S. D.	4.03	5.14	9.90	2.33	3.33	5.47	3.52	4.13	7.55
Total									
Mean	14.0	13.8	27.8	13.5	13.1	26.6	13.8	13.2	27.0
S. D.	6.57	5.83	12.18	4.02	4.00	5.88	4.52	4.28	8.61

TABLE IV

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
BREATHING DEPTH-VARIANCE MEANS

	High Verbal Group			Equivalent Group			High Math Group			Total Group Means
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	
Initial Testing										
Male Mean	52.5	38.3	90.8	29.3	23.6	52.9	28.9	26.0	55.0	66.2
S.D.	45.23	25.33	22.17	15.34	16.37	31.23	19.98	15.05	34.91	51.70
Female Mean	14.5	19.3	33.8	19.7	20.3	40.0	17.2	14.7	31.9	35.3
S.D.	13.84	17.45	30.70	10.42	10.98	20.80	8.78	7.35	15.23	23.47
Total Mean	33.5	28.8	62.3	24.5	21.9	46.5	23.1	20.4	43.5	50.7
S.D.	38.48	23.72	61.17	13.97	12.77	27.31	16.51	13.13	29.37	43.03
Final Testing										
Male Mean	51.5	38.4	90.0	28.7	27.7	56.3	25.8	23.2	49.0	65.1
S.D.	38.92	26.37	64.87	20.96	16.14	36.53	14.97	12.90	27.48	19.20
Female Mean	14.3	17.3	31.6	18.0	17.1	35.2	33.1	20.9	43.0	36.6
S.D.	12.40	15.88	27.75	6.27	10.05	14.81	9.78	12.08	21.22	22.65
Total Mean	32.9	27.8	60.8	23.4	22.4	45.7	23.9	22.1	46.0	50.8
S.D.	34.36	24.20	57.80	19.62	14.44	30.14	12.77	12.55	24.73	40.86

TABLE V

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
BREATHING RATE-MEAN MEANS

	High Verbal Group			Equivalent Group			High Math Group			Total Group Means
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	
Initial Testing										
Male										
Mean	1.6	1.6	3.2	1.6	1.7	3.3	1.7	1.7	3.4	3.3
S.D.	.37	.37	.73	.32	.28	.57	.01	.01	.19	.55
Female										
Mean	1.7	1.7	3.4	1.6	1.6	3.2	1.9	1.8	3.6	3.4
S.D.	.36	.37	.72	.18	.20	.37	.31	.28	.58	.60
Total										
Mean	1.6	1.7	3.3	1.6	1.7	3.3	1.8	1.7	3.5	3.4
S.D.	.37	.37	.73	.26	.25	.49	.25	.21	.46	.58
Final Testing										
Male										
Mean	1.5	1.6	3.1	1.6	1.6	3.3	1.6	1.6	3.2	3.2
S.D.	.45	.37	.82	.10	.23	.18	.28	.22	.49	.57
Female										
Mean	1.7	1.6	3.3	1.9	1.7	3.6	1.5	1.6	3.2	3.3
S.D.	.25	.36	.59	.30	.24	.53	.17	.18	.34	.52
Total										
Mean	1.6	1.6	3.2	1.7	1.7	3.4	1.6	1.6	3.7	3.3
S.D.	.38	.37	.72	.25	.23	.42	.23	.20	.42	.55

TABLE VI

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
BREATHING RATE-VARIANCE MEANS

	High Verbal Group			Equivalent Group			High Math Group			Total Group Means
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	
					Initial Testing					
Male Mean	.3	.3	.7	.3	.3	.7	.3	.4	.7	.7
S. D.	.15	.10	.24	.12	.09	.20	.13	.05	.09	.19
Female Mean	.4	.4	.7	.3	.3	.6	.4	.4	.8	.7
S. D.	.14	.13	.26	.06	.07	.11	.14	.12	.25	.23
Total Mean	.3	.4	.7	.3	.3	.7	.4	.4	.7	.7
S. D.	.14	.12	.25	.09	.08	.16	.11	.09	.20	.22
					Final Testing					
Male Mean	.3	.3	.6	.4	.3	.7	.3	.3	.7	.7
S. D.	.18	.15	.32	.05	.04	.08	.11	.09	.19	.22
Female Mean	.4	.3	.7	.4	.4	.8	.3	.4	.7	.7
S. D.	.10	.14	.23	.15	.13	.22	.05	.07	.11	.23
Total Mean	.3	.3	.7	.4	.4	.7	.3	.3	.7	.7
S. D.	.15	.14	.28	.11	.07	.18	.09	.08	.16	.21

TABLE VII

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
BLOOD PRESSURE-MEAN MEANS

	High Verbal Group			Equivalent Group			High Math Group			Total Group Means
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	
Initial Testing										
Male Mean	8.2	8.6	16.9	9.1	9.3	18.4	8.5	9.5	18.0	17.8
S.D.	.99	1.22	2.05	.75	1.46	1.87	1.61	1.23	2.54	2.27
Female Mean	8.3	7.6	15.9	7.7	7.5	15.1	9.5	9.2	18.8	16.6
S.D.	1.46	1.10	2.36	1.12	1.35	2.10	1.78	.77	2.37	2.79
Total Mean	8.3	8.1	16.4	8.4	8.4	16.8	9.0	9.4	18.4	17.2
S.D.	1.25	1.27	2.27	1.20	1.68	2.59	1.77	1.04	2.48	2.60
Final Testing										
Male Mean	10.9	9.6	20.5	10.1	10.7	20.8	11.3	10.6	21.9	21.0
S.D.	2.09	2.17	3.87	2.64	1.63	4.09	2.06	1.67	3.57	3.89
Female Mean	10.4	8.2	18.6	10.5	10.9	21.4	8.7	7.8	16.5	18.8
S.D.	1.89	.84	2.57	1.48	2.65	3.86	2.23	1.44	3.47	3.89
Total Mean	10.6	7.9	19.6	10.3	10.8	21.1	10.0	9.2	19.2	19.9
S.D.	2.00	1.78	3.41	2.15	2.20	3.98	2.53	2.07	4.43	4.04

TABLE VIII

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
BLOOD PRESSURE-VARIANCE MEANS

	High Verbal Group			Equivalent Group			High Math Group			Total Group Means
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	
Initial Testing										
Male										
Mean	7.6	8.5	16.2	9.4	10.0	19.4	8.7	10.3	19.1	18.2
S.D.	1.83	2.38	3.97	1.43	3.11	3.77	3.14	2.67	5.25	4.62
Female										
Mean	7.9	6.6	14.5	6.9	6.5	13.4	10.5	9.5	20.0	16.0
S.D.	2.76	1.82	4.33	1.85	2.06	3.38	3.59	1.56	4.80	5.10
Total										
Mean	7.8	7.6	15.3	8.2	8.2	16.4	9.6	9.9	19.5	17.1
S.D.	2.37	2.33	4.23	2.07	3.17	4.68	3.49	2.22	5.05	4.99
Final Testing										
Male										
Mean	14.0	11.0	25.0	12.5	13.1	25.6	15.0	12.8	27.8	26.1
S.D.	5.53	5.17	9.95	5.90	3.98	14.16	5.45	3.98	9.08	9.57
Female										
Mean	12.6	7.6	20.2	12.8	14.2	27.0	9.5	7.3	16.8	21.3
S.D.	4.58	1.53	5.86	3.30	6.19	8.73	4.44	2.49	6.39	8.30
Total										
Mean	13.3	9.3	22.6	12.6	13.6	26.3	12.2	10.0	22.3	23.7
S.D.	5.12	4.17	8.50	4.78	5.23	9.12	5.62	4.29	9.53	9.25

TABLE IX

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
HEART RATE-MEAN MEANS

	High Verbal Group			Equivalent Group			High Math Group			Total Group Means
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	
					Initial Testing					
Male										
Mean	7.2	7.5	14.6	7.2	7.2	14.4	6.6	6.5	13.1	14.0
S.D.	.93	.99	1.91	.88	1.14	2.00	1.15	1.13	2.24	2.06
Female										
Mean	6.8	6.9	13.7	7.5	7.7	15.2	7.62	7.5	15.1	14.7
S.D.	.78	.93	1.69	.90	.93	1.82	1.06	1.04	2.05	1.99
Total										
Mean	7.0	7.2	14.1	7.3	7.5	14.8	7.1	7.0	14.1	14.3
S.D.	1.02	.98	1.99	.83	1.26	2.08	1.01	.97	1.95	2.07
					Final Testing					
Male										
Mean	7.2	7.3	14.5	7.2	7.0	14.2	6.8	6.6	13.3	14.0
S.D.	1.34	.99	1.95	.93	1.06	1.99	1.15	1.04	2.17	2.09
Female										
Mean	6.8	6.7	13.5	7.7	7.5	15.2	7.7	7.6	15.3	14.7
S.D.	.79	.83	1.59	.75	.82	1.54	1.54	.98	1.96	1.90
Total										
Mean	7.0	7.0	14.0	7.4	7.3	14.7	7.2	7.1	14.3	14.3
S.D.	.91	.96	1.85	.89	.98	1.86	1.19	1.12	2.28	2.03

TABLE X

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
HEART RATE-VARIANCE MEANS

	High Verbal Group			Equivalent Group			High Math Group			Total Group Means
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	
Initial Testing										
Male										
Mean	5.8	6.2	12.0	5.7	5.9	11.6	4.8	4.7	9.5	11.0
S.D.	1.55	1.50	3.03	1.32	1.94	3.23	1.28	1.33	2.59	3.15
Female										
Mean	5.1	5.3	10.4	6.3	6.6	13.0	6.5	6.3	12.8	12.1
S.D.	1.15	1.43	2.56	1.50	1.59	3.08	1.82	1.68	3.41	3.25
Total										
Mean	5.4	5.8	11.2	6.0	6.3	12.3	5.7	5.5	11.2	11.6
S.D.	1.40	1.53	2.91	1.45	1.81	3.23	1.79	1.72	3.45	3.25
Final Testing										
Male										
Mean	5.8	6.0	11.7	5.8	5.5	11.3	5.2	4.9	10.1	11.0
S.D.	1.48	1.57	3.02	1.51	1.52	3.18	1.64	1.48	3.06	3.16
Female										
Mean	5.1	5.0	10.2	6.7	6.3	13.0	6.6	6.4	13.0	12.0
S.D.	1.17	1.21	2.35	1.26	1.38	2.59	1.74	1.60	3.25	3.06
Total										
Mean	5.4	5.5	10.9	6.2	5.9	12.1	5.9	5.6	11.5	11.5
S.D.	1.37	1.48	2.82	1.46	1.51	3.01	1.84	1.70	3.47	3.15

TABLE XI

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
P G R - MEAN MEANS

	High Verbal Group			Equivalent Group			High Math Group			Total Group Means
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	
					Initial Testing					
Male										
Mean	19.1	20.4	39.5	15.2	17.7	32.9	14.3	14.0	28.3	33.5
S.D.	3.52	2.32	5.59	5.64	4.02	9.21	3.21	2.45	4.07	8.09
Female										
Mean	14.6	15.3	29.9	17.3	18.8	36.1	18.5	16.5	35.0	33.7
S.D.	4.19	4.71	8.70	3.16	3.21	5.99	2.34	2.37	4.55	7.17
Total										
Mean	16.9	17.8	34.7	16.3	18.3	34.5	16.4	15.2	31.6	33.6
S.D.	4.46	4.51	8.75	4.68	3.68	7.94	3.52	2.72	5.49	7.65
					Final Testing					
Male										
Mean	19.5	20.0	39.6	16.4	17.1	33.5	14.9	13.0	27.9	33.7
S.D.	3.79	2.42	5.80	4.70	3.92	7.26	2.90	1.97	4.41	7.60
Female										
Mean	14.2	15.2	29.5	16.5	18.7	35.2	17.7	16.2	33.9	32.8
S.D.	4.12	5.00	9.00	3.89	3.50	6.97	2.39	2.25	4.47	7.47
Total										
Mean	16.9	17.6	34.5	16.4	17.9	34.3	16.3	14.6	30.9	33.2
S.D.	4.76	4.60	9.10	4.31	3.84	7.16	3.00	2.64	5.34	7.55

TABLE XII

MEANS AND STANDARD DEVIATIONS BY GROUP, SEXES, AND TESTS
P G R-VARIANCE MEANS

	High Verbal Group			Equivalent Group			High Math Group			Total Group Means
	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	ACE Math	ACE Verbal	Total	
Initial Testing										
Male Mean	41.5	46.4	87.9	29.5	36.2	65.7	24.2	22.3	46.5	66.7
S.D.	15.12	10.31	24.28	17.27	13.98	28.84	9.66	7.88	12.79	28.02
Female Mean	25.6	28.7	54.3	34.4	40.0	74.4	39.3	30.6	69.9	66.2
S.D.	13.77	17.34	30.60	11.25	13.78	23.95	10.54	9.28	18.81	26.38
Total Mean	33.6	37.5	71.1	32.0	38.1	70.1	31.7	26.4	58.2	66.4
S.D.	16.5	16.79	32.34	14.78	14.00	26.86	12.63	9.55	19.89	27.49
Final Testing										
Male Mean	43.6	44.8	88.4	32.1	33.8	65.9	25.7	19.5	45.2	66.5
S.D.	17.30	10.61	25.85	19.5	13.28	27.78	9.34	5.79	13.81	29.24
Female Mean	24.1	28.3	52.4	31.5	40.0	71.5	34.8	29.4	64.2	62.7
S.D.	13.09	18.57	31.22	12.85	15.38	26.52	9.93	8.57	17.98	27.00
Total Mean	33.8	36.6	70.4	31.8	36.9	68.7	30.3	24.4	54.7	54.0
S.D.	18.18	17.22	33.85	16.55	14.70	27.30	10.67	8.83	18.63	28.2

TABLE XIII

ANALYSIS OF VARIANCE FOR BREATHING DEPTH
MEAN MEANS ON INITIAL TESTING

Source of Variable	Sum of Squares	df	Mean Square	F Ratio
Between Groups	33.448	2	16.724	<1
Between Sexes	424.2	1	424.200	8.55**
Group × Sex	131.531	2	65.766	1.326
Between Sexes in G × S	2083.206	42	49.600	
	2672.385	47		
M vs V	9.375	1	9.375	4.137*
(M - V) × Group	3.425	2	1.713	<1
(M - V) × Sex	18.551	1	18.551	8.187**
(M - V) Sex × Group	5.67	2	2.835	1.251
Residual (error)	95.179	42	2.266	
Total within Sexes	132.2	48		
Total	2804.585	95		

* Indicates significance at .05 level. ** Indicates significance at .01 level.

TABLE XIV

ANALYSIS OF VARIANCE FOR BREATHING DEPTH
MEAN MEANS ON FINAL TESTING

Source of Variable	Sum of Squares	df	Mean Square	F Ratio
Between Groups	6.145	2	3.073	<1
Between Sexes	315.737	1	315.737	7.245
Group × Sex	136.117	2	68.059	1.56
Between Sexes in G × S	1830.328	42	43.579	
Total between Sexes	2288.327	47		
M vs V	4.815	1	4.815	2.767
(M - V) × Group	.498	2	.248	<1
(M - V) × Sex	3.118	1	3.118	1.79
(M - V) × Sex × Group	11.339	2	5.67	3.259*
Residual (error)	73.088	42	1.74	
Total within Sexes	92.856	48		
Total	2381.183	95		

* Indicates significance at .05 level.

TABLE XV

ANALYSIS OF VARIANCE FOR *BREATHING DEPTH*
VARIANCE MEANS ON INITIAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	1642.9	2	821.45	<1
Between Sexes	5756.7	1	5756.7	6.827*
Groups × Sex	2134.4	2	1067.2	1.284
Between Sexes in G × S	34904.1	42	831.05	
Total between Sexes	44438.1	47		
M vs V	268.7	1	268.7	4.54
(M - V) × Group	23.2	2	11.6	<1
(M - V) × Sex	545.5	1	545.5	9.22
(M - V) × Sex × Group	265.2	2	132.6	2.24
Residual (error)	2484.8	42	59.162	
Total within Sexes	3587.4	48		
Total	48025.5	95		

* Indicates significance at the .05 level.

TABLE XVI

ANALYSIS OF VARIANCE FOR BREATHING DEPTH
VARIANCE MEANS ON FINAL TESTING

Source of Variable	Sum of Squares	df	Mean Square	F Ratio
Between Groups	1182.343	2	591.172	<1
Between Sexes	4870.65	1	4870.65	6.794*
Group × Sex	2910.842	2	1455.421	2.03
Between Sexes in G × S	30108.951	42	716.88	
Total between Sexes	39072.786	47		
M vs V	167.484	1	167.484	2.188
(M - V) × Group	75.849	2	37.925	<1
(M - V) × Sex	205.332	1	205.332	2.683
(M - V) × Sex × Group	313.624	2	156.812	2.049
Residual (error)	3214.641	42	76.539	
Total within Sexes	3976.93	48		
Total	43049.716	95		

* Indicates significance at the .05 level.

TABLE XVII

ANALYSIS OF VARIANCE FOR BREATHING RATE
MEAN MEANS ON INITIAL TESTING

Source of Variable	Sum of Squares	df	Mean Square	F Ratio
Between Groups	.232	2	.116	<1
Between Sexes	.088	1	.088	<1
Group × Sex	.173	2	.087	<1
Between Sexes in G × S	7.556	42	.18	
Total between Sexes	8.049	47		
M vs V	.005	1	.005	<1
(M - V) × Group	.041	2	.021	2.63
(M - V) × Sex	.005	1	.005	<1
(M - V) × Sex × Group	.011	2	.011	1.38
Residual (error)	.333	42	.008	
Total within Sexes	.395	48		
Total	8.444	95		

TABLE XVIII

ANALYSIS OF VARIANCE FOR BREATHING RATE
MEAN MEANS ON FINAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	.224	2	.112	<1
Between Sexes	.107	1	.107	<1
Group × Sex	.14	2	.07	<1
Between Sexes in G × S	6.739	42	.16	
Total between Sexes	7.21	47		
M vs V	.007	1	.007	<1
(M - V) × Group	.069	2	.035	3.89*
(M - V) × Sex	.05	1	.05	5.56*
(M - V) × Sex × Group	.063	2	.032	3.56*
Residual (error)	.361	42	.009	
Total within Sexes	.55	48		
Total	7.76	95		

* Indicates significance at the .05 level.

TABLE XIX

ANALYSIS OF VARIANCE FOR BREATHING RATE
VARIANCE MEANS ON INITIAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	.037	2	.019	<1
Between Sexes	.018	1	.018	<1
Group $\times$ Sex	.018	2	.009	<1
Between Sexes in G $\times$ S	1.011	42	.024	
Total between Sexes	1.084	47		
M vs V	.001	1	.001	<1
(M - V) $\times$ Group	.007	2	.004	<1
(M - V) $\times$ Sex	.002	1	.002	<1
(M - V) $\times$ Sex $\times$ Group	.007	2	.004	<1
Residual (error)	.078	42	.019	
Total within Sexes	.095	48		
Total	1.179	95		

TABLE XX

ANALYSIS OF VARIANCE FOR BREATHING RATE
VARIANCE MEANS ON FINAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	.03	2	.015	<1
Between Sexes	.02	1	.02	<1
Group $\times$ Sex	.01	2	.005	<1
Between Sexes in G $\times$ S	1.04	42	.025	
Total between Sexes	1.1	47		
M vs V	.0	1	0	<1
(M - V) $\times$ Group	.02	2	.01	<1
(M - V) $\times$ Sex	.0	1	0	<1
(M - V) $\times$ Sex $\times$ Group	.01	2	.005	<1
Residual (error)	110.8	42	2.638	
Total within Sexes	110.83	48		
Total	111.93	95		

TABLE XXI

ANALYSIS OF VARIANCE FOR BLOOD PRESSURE
MEAN MEANS ON INITIAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	18.6	2	9.300	3.21*
Between Sexes	8.6	1	8.600	3.03
Group × Sex	16.3	2	8.150	2.87
Between Sexes in G and S	119.2	42	2.8438	
Total between Sexes	162.7	47		
M v V	.2	1	.200	<1
(M - V) × Group	.9	2	.900	<1
(M - V) × Sex	5.0	1	5.000	5.780*
(M - V) × Sex × Group	1.1	2	.550	<1
Residual (error)	36.3	42	.864	
Total within Sexes	43.5	48		
Total	206.2	95		

* Indicates significance at the .05 level.

TABLE XXII

ANALYSIS OF VARIANCE FOR BLOOD PRESSURE
MEAN MEANS ON FINAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	15.415	2	7.708	1.039
Between Sexes	29.261	1	29.261	3.947
Group × Sex	36.384	2	18.192	2.952
Between Sexes in G × S	311.309	42	7.412	
Total between Sexes	392.369	47		
M vs V	11.761	1	11.761	8.436**
(M - V) × Group	18.983	2	9.492	6.809**
(M - V) × Sex	1.0	1	1.0	<1
(M - V) × Sex × Group	.66	2	.33	<1
Residual (error)	58.546	42	1.394	
Total within Sexes	90.95	48		
Total	483.319	95		

** Indicates significance at the .01 level.

TABLE XXIII

ANALYSIS OF VARIANCE FOR BLOOD PRESSURE
VARIANCE MEANS ON INITIAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	75.4	2	37.7	3.57*
Between Sexes	30.3	1	30.3	2.87
Group × Sex	49.2	2	24.6	2.33
Between Sexes in G × S	443.1	42	10.55	
Total between Sexes	598.0	47		
M vs V	.1	1	.1	< 1
(M - V) × Group	1.2	2	.6	< 1
(M - V) × Sex	22.5	1	22.5	7.18*
(M - V) × Sex × Group	2.7	2	1.35	< 1
Residual (error)	131.6	42	3.133	
Total within Sexes	158.1	48		
Total	756.1	95		

* Indicates significance at the .05 level.

TABLE XXIV

ANALYSIS OF VARIANCE FOR BLOOD PRESSURE
VARIANCE MEANS ON FINAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	78.5	2	39.25	< 1
Between Sexes	137.4	1	137.5	3.42
Group × Sex	153.2	2	76.6	1.908
Between Sexes in G × S	1686.4	42	40.152	
Total between Sexes	2055.6	47		
M vs V	73.0	1	73.0	9.01**
(M - V) × Group	101.7	2	50.85	6.276*
(M - V) × Sex	.7	1	.7	< 1
(M - V) × Sex × Group	8.6	2	4.3	< 1
Residual (error)	340.3	42	8.102	
Total within Sexes	524.3	48		
Total	2579.9	95		

* Indicates significance at .05 level.

** Indicates significance at .01 level.

TABLE XXV

ANALYSIS OF VARIANCE FOR HEART RATE
MEAN MEANS ON INITIAL TESTING

Source of Variable	Sum of Squares	df	Mean Square	F Ratio
Between Groups	2.5	2	1.25	<1
Between Sexes	2.32	1	2.32	1.09
Group × Sex	9.242	2	4.621	2.165
Between Sexes in G × S	89.643	42	2.134	
Total between Sexes	103.705	47		
M vs V	.133	1	.133	1.82
(M - V) × Group	.518	2	.259	3.548*
(M - V) × Sex	.055	1	.055	<1
(M - V) × Sex × Group	.048	2	.024	<1
Residual (error)	3.051	42	.073	
Total within Sexes	3.805	48		
Total	107.51	95		

* Indicates significance at the .05 level.

TABLE XXVI

ANALYSIS OF VARIANCE FOR HEART RATE
MEAN MEANS ON FINAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	2.181	2	1.091	<1
Between Sexes	2.7	1	2.7	1.34
Group × Sex	8.931	2	4.466	2.21
Between Sexes in G × S	84.768	42	2.018	
Total between Sexes	98.58	47		
M vs V	.25	1	.25	4.24*
(M - V) × Group	.215	2	.108	1.83
(M - V) × Sex	.2024	1	.024	<1
(M - V) × Sex × Group	.048	2	.024	<1
Residual (error)	2.458	42	.059	
Total within Sexes	2.995	48		
Total	101.575	95		

* Indicates significance at the .05 level.

TABLE XXVII

ANALYSIS OF VARIANCE FOR HEART RATE VARIANCE MEANS ON INITIAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	6.421	2	3.211	<1
Between Sexes	6.51	1	6.51	1.266
Groups × Sex	23.953	2	11.977	2.33
Between Sexes in G × S	215.936	42	5.141	
Total between Sexes	252.82	47		
M vs V	.51	1	.51	3.09
(M - V) × Group	1.403	2	.702	4.23*
(M - V) × Sex	.043	1	.043	<1
(M - V) × Sex × Group	.045	2	.023	<1
Residual (error)	6.919	42	.165	
Total within Sexes	8.92	48		
Total	261.74	95		

* Indicates significance at the .05 level.

TABLE XXVIII

ANALYSIS OF VARIANCE FOR HEART RATE VARIANCE MEANS ON FINAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	5.6	2	2.8	<1
Between Sexes	6.1	1	6.1	1.249
Groups × Sex	21.4	2	10.7	2.191
Between Sexes in G × S	205.1	42	4.883	
Total between Sexes	238.2	47		
M vs V	.7	1	.7	4.516*
(M - V) × Group	.6	2	.3	1.935
(M - V) × Sex	0	1	0	<1
(M - V) × Sex × Group	.2	2	.1	<1
Residual (error)	6.5	42	.155	
Total within Sexes	8.0	48		
Total	246.2	95		

* Indicates significance at the .05 level.

TABLE XXIX

ANALYSIS OF VARIANCE FOR P G R
MEAN MEANS ON INITIAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	47.2	2	23.6	<1
Between Sexes	.2	1	.2	<1
Group × Sex	296.1	2	148.05	5.866**
Between Sexes in G × S	1060.0	42	25.238	
Total between Sexes	1403.5	47		
M vs V	8.2	1	8.2	2.08
(M - V) × Group	41.3	2	20.65	5.24**
(M - V) × Sex	7.6	1	7.6	1.93
(M - V) × Sex × Group	1.3	2	.65	<1
Residual (error)	165.5	42	3.94	
Total within Sexes	223.9	48		
Total	1627.4	95		

** Indicates significance at the .01 level.

TABLE XXX

ANALYSIS OF VARIANCE FOR P G R
MEAN MEANS ON FINAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	6.284	2	33.142	1.363
Between Sexes	4.208	1	4.208	<1
Groups × Sex	275.551	2	137.776	5.666**
Between Sexes in G × S	1021.272	42	24.316	
Total between Sexes	1367.315	47		
M vs V	.789	1	.789	<1
(M - V) × Group	43.136	2	21.568	4.875*
(M - V) × Sex	3.880	1	3.880	<1
(M - V) × Sex × Group	1.516	2	.758	<1
Residual (error)	185.064	42	4.406	
Total within Sexes	234.385	48		
Total	1601.700	95		

* Indicates significance at the .05 level. ** Indicates significance at the .01 level.

TABLE XXXI

ANALYSIS OF VARIANCE FOR P G R
VARIANCE MEANS ON INITIAL TESTING

Source of Variable	Sum of Square	df	Mean Square	F Ratio
Between Groups	826.84	2	413.42	1.26
Between Sexes	1.65	1	1.65	<1
Group × Sex	3512.86	2	1756.43	5.349**
Between Sexes in G × S	13789.6	42	328.324	
Total between Sexes	18130.95	47		
M vs V	62.4	1	62.4	1.239
(M - V) × Group	589.81	2	294.905	5.858**
(M - V) × Sex	63.7	1	63.7	1.265
(M - V) × Sex × Group	39.55	2	19.775	<1
Residual (error)	2114.33	42	50.341	
Total within Sexes	2869.79	48		
Total	21000.74	95		

** Indicates significance at the .01 level.

TABLE XXXII

ANALYSIS OF VARIANCE FOR P G R
VARIANCE MEANS ON FINAL TESTING

Source of Variable	Sum of Squares	df	Mean Square	F Ratio
Between Groups	1191.2	2	595.6	1.721
Between Sexes	86.5	1	86.5	<1
Group × Sex	3287.8	2	1643.9	4.75*
Between Sexes in G × S	14528.3	42	345.912	
Total between Sexes	19093.8	47		
M vs V	11.0	1	11.0	<1
(M - V) × Group	528.4	2	264.2	3.799*
(M - V) × Sex	74.4	1	74.4	1.07
(M - V) × Sex × Group	41.8	2	20.9	<1
Residual (error)	2920.2	42	69.529	
Total within Sexes	3575.8	48		
Total	22669.6	95		

* Indicates significance at the .05 level.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 1, September 1962)

THE ACCURACY OF SELF-ROLE PERCEPTION: A DEVELOPMENTAL STUDY

JOHN E. deJUNG
University of Kansas
ERIC F. GARDNER
Syracuse University

ACCURACY OF interpersonal perception has been variously studied, using sociometric nominational procedures, personality and attitudinal ratings, projective techniques, and varying other approaches. 1* Relatively few of these studies have examined this accuracy-inaccuracy continuum developmentally. The present paper introduces sociometric procedure providing a measure of this accuracy variable in school children. This procedure is an extension of a promising new technique for measuring intermember perception of social need satisfaction yielding both individual and group indices (10). The accuracy measure, defined as the average absolute difference between a pupil's expected rating from another and the actual rating he receives from that other, is examined in terms of its relationships with other rating indices. Certain basic developmental hypotheses, suggestive of a postulated physio-social maturation process underlying social perceptual behavior, are examined and discussed.

Three earlier studies lend pertinence to the present report. Dymond, Hughes, and Raabe in 1952 (7) reported differences in perceptive ability as measured by emphatic responses to a modified TAT and as measured by agreement of the subject's estimated rating of how others feel (like-dislike) toward the subject, and the subject's sociometric score based on number of times chosen by others in three social situations (sit next to, invite to a party, transfer to another room with). Scores for 24 second graders and 16 sixth graders were compared on both empathetic measures and significant differences favoring the older group were reported.

Ausubel, Schiff, and Gasser (2) studied trends in sociempathy defined in terms of perception of own status (status was measured as a consensus of others' friendship ratings along a five point scale of desirability) involving a grade school and a university high school sample of third, fifth, seventh, eleventh, and twelfth graders. A correlation coefficient, relating the means of each individual's several predictions of how others rate him with the mean of his several received ratings was obtained for each of the several grades and sex subgroups within grades. Relatively consistent increasing trends in sociempathetic ability with age were found, with the greatest gain between grades eleven and twelve. The data further suggested that boys were more perceptive of ratings received from opposite sex members than of those received from same sex members. The data for girls tended to evince an opposite picture (girls being more aware of other girls' ratings), though reversals were prominent here. In a later report, Ausubel and Schiff (1) describe the greater apparency of sociometric attitudes held by eleventh grade girls than by eleventh grade boys.

Lundberg and Dickson (13) examined accuracy in predicting sociometric reciprocation as part of a study of interethnic relations in a large high school (N = 1500). Their data were based on student nominations of their three best friends and the prediction of which of these nominated friends had also chosen them. The authors report that accuracy in prediction of reciprocity apparently increases with age and school class, approaching a peak in the junior year and falling back somewhat in the senior year.

Though these studies provide a general description of social awareness in the pupil population, the differing methodologies do not permit specific interstudy support (or contradiction). The lack of more precise formulation of the underlying continuum (general friendship bond), itself embodying varying meanings with age, reduces interpretability of intergrade differences. Further experimental limitations are characteristic of certain of these exploratory studies: the problems 1) of a very few key persons in a group possibly distorting the responses received by less extreme individuals, 2) of the ambiguous meanings of omission, 3) of probability of choice as a

*Footnotes will be found at the end of the article.

function of recall, 4) of rater tendency to avoid nominating their peers on adverse social criteria. Tagiuri, Blake, and Bruner stress several of these problems in their discussion of conditions associated with accuracy of self-perception and emphasize the need for clinical data and longitudinal study (18).

This present study is offered as a further examination of pupil behavior in varying age and sex subgroups with respect to their awareness of perceptions of themselves by their classmates. An explicit examination of the hypotheses examined in this study is presented following a description of the sample tested and an introduction and examination of the perceptual accuracy index employed.

Subjects and Testing Procedure

A public school sample consisting of the entire attending pupil population of grades five through twelve in a moderately sized central school in upper New York State was tested in this study. The school district may be described as one composed of relatively homogeneous population of merchant, farmer, and laborer families with minimal obvious social class distinctions. All pupils, with a very minor percentage of transferrals, had been housed in the single school building for their entire school career, thus minimizing the kind of social problems introduced by graduation from elementary to junior or senior high school buildings or by the sudden widening of pupil populations as is the case when a "country" school graduates its pupils into the larger town school building. A general policy of social promotion prevailed in the school.

In terms of internal school organization, there were two "home room" classes at each grade level. The intra-grade division at the fifth and sixth grade levels was random; at the junior high, "loosely" by academic achievement; and at senior high, alphabetically. To maintain administratively efficient sized groups (20-40) and to minimize intellective or family name bias within subgroups, each junior and senior high school grade was randomly divided into two testing sections. The fifth and sixth grade sections were left "intact." The socio-economic homogeneity of the town and pupil population, the single school building system, and the relatively small student enrollment, all promoted a high degree of interpersonal recognition and contact among pupils, especially within any single grade level.

The testing was accomplished in a single day at the end of the sixth month of the school year by a team of advanced graduate psychology students, two grade levels tested at each hour. Each pupil received a scale booklet with a list of names of the pupils in his section. These lists varied in number from 18 t0 34. Three hundred and eighty-seven pupils were tested; 21 pupils were either absent or not in the rooms at the time of the scale administration; two test booklets were later discarded because of obvious lack of respondent co-operation. Ratings of absentee pupils were also discarded. All pupil rating indices used in this study are averages based on the number of ratings made or grade means based on the average of these pupil averages. Age and IQ data were obtained from the school record file. Except for the few transfer students, all of the intelligence quotients were based on the California Tests of Mental Maturity administered within two years of the present study. Table I describes the obtained distributions of these variables, the average number of usable ratings made by the pupils, and the number of pupils enrolled at each grade level.

The Syracuse Scales

Basic to the measure of accuracy of social perception introduced in this study is the theory and format of the Syracuse Scales of Social Relations designed by Gardner and Thompson (10). The instrument provides a scale permitting comparison of groups or individuals in terms of their potential for satisfying psychologically defined needs. The initial research leading to this scale involved extended testing and retesting of college fraternity groups. More recently, these scales have been revised downwards and incorporated into several research studies involving elementary and secondary school populations (9, 11).

The Syracuse Scales are administered by the classroom teacher, as are most standardized tests, by the aid of directions which are read verbatim. The pupil is presented with a matrix of rectangles defined vertically by the names of all his classmates and horizontally by his reference population. The pupil is first directed (by clearly worded instructions and supervision) to describe his reference population by writing, first, the name of the person of all the persons he has ever known that he would most like to have help him or be with him (in reference to some well-defined need situation), then the individual least liked, and so on, following a fractionation process involving five names. These names, after selection by the pupil from his population of all the persons he has ever known, are entered across the top of his page in rank sequence. This procedure is designed to set up a reference continuum, approximating an equal interval scale, against which the pupil will now compare and locate each of his classmates. The appropriateness of this procedure has been supported by empirical demonstrations (10, chapter 15).

TABLE I

DISTRIBUTIONS OF SEX, CA, IQ (AS DETERMINED FROM THE CALIFORNIA TESTS OF MENTAL MATURITY), AVERAGE NUMBER OF RATINGS MADE PER PUPIL, AND NUMBER OF PUPILS PER GRADE FOR THE PUBLIC SCHOOL SAMPLE, GRADES 5-12

Grade	5	6	7	8	9	10	11	12	Total
Sex									
Girls	24	34	37	26	23	26	21	19	210
Boys	31	22	27	26	27	10	14	18	175
Total	55	56	64	52	50	36	35	37	385
CA (mos.)									
$\bar{x}$	131	142	156	170	178	192	202	216	169
σ	8.8	9.2	10.5	10.7	8.0	7.5	6.9	7.3	28.5
IQ									
$\bar{x}$	116	107	106	108	104	99	106	106	107
σ	16.3	13.3	14.6	14.3	9.6	8.8	8.8	8.2	13.4
Average No. of usable ratings made per pupil	26	27	31	25	24	17	16	17	24
Pupil enrollment	56	57	64	57	57	38	37	42	408

TABLE II

WITHIN-GRADE PRODUCT MOMENT CORRELATIONS OF DERIVED ESTIMATIVE INDICES AND MEAN ABSOLUTE Dg SCORE WITH DERIVED RATING INDICES

Correlates	Grade 5 / N 55	6 / 56	7 / 64	8 / 52	9 / 50	10 / 36	11 / 35	12 / 37	Average / 385
1. Mean Estimate Made Mean Rating Received	-.23	.05	-.06	.36	.18	.16	.15	-.15	.08
2. $\lvert \bar{Dg} \rvert$ and Mean Rating Received	.05	.16	.38	.51	.28	.21	.52	.09	.27
3. $\lvert \bar{Dg} \rvert$ and σ^2 of Ratings Received	.04	.01	.41	.49	.15	.16	.35	.04	.21
4. $\lvert \bar{Dg} \rvert$ and Absolute Deviation of Mean Rating Received from Group Mean Rating	.46	.20	.34	.24	.33	.09	.14	.15	.27
5. $\lvert \bar{Dg} \rvert$ and Absolute Deviation of Mean Estimate made from Group Mean Rating Received	.39	.02	.45	.02	.32	.04	.00	.37	.23

Both the development of the reference population and the pupil comparisons are confined to consideration of one need situation at the time. The "need situation" chosen for this study was need-succorance, defined in the Syracuse Scales by the following reference situation:

"Sometimes you get into trouble and you feel unhappy. It might be that you had been blamed for something you didn't do. Think about some time when you were very unhappy and would have liked to talk over your troubles with some kind, sympathetic person."

Selection of this particular need was made in terms of meaningfulness to pupil, adaptableness to the testing situation, stability of personal relevance over differing age and social groups, probable freedom from prior teacher or group pronouncement, and relative lack of response-inhibiting associations.

Two basic indices obtainable from the Syracuse Scales, individual mean ratings made and individual mean ratings received, are measures of each group member's consensual appraisal of his group's potential for satisfying certain of his psychological needs as defined by the reference situation, and measures of the group's consensual appraisal of himself in turn (again with respect to this same need satisfaction potential). The group mean for either set of these individual means may be interpreted in terms of average group appraisal made (or received; these two group averages are arithmetically identical). In both logic of design and obtained indices, the Syracuse Scales format suggests a valuable advance over prior sociometric procedures in; 1) reducing ambiguity of choice bases by explicit definition of the choice-continuum; 2) utilizing reference populations which psychologically extend beyond the immediate group permitting inter-group comparisons; 3) providing ratings on all group members by all other group members; and 4) providing approximal equal interval scaling properties and obviating the tendency to equate first and second choices and dichotomize choice versus non-choice response.

The Dg Score

In examining accuracy of self-perception, the format of the Syracuse Scales is extended to include an anticipated rating. The Syracuse Scales is first administered and each pupil receives a rating with respect to a specifically defined need situation from each of the classmates in his rating section. He is then asked to rate himself on the same need situation as he thinks each of his classmates has rated him. He is required to estimate only for those classmates who have just rated him. He is cautioned to be as accurate as he is able in making these estimates, that the administrator is primarily interested "in seeing how close he can come." These two sets of ratings, the received and the guessed, furnish basic data for a series of discrepancy scores, several of which suggest pertinence to the study of group member interaction.

The social perception accuracy index used in this study is a distance function, Dg, defined as the difference between the rating a pupil receives on the Syracuse Scales and the rating he expects or guesses he will receive. Each pupil will have as many discrepancy scores as there are pupils rating him. His average absolute Dg score is that pupil's index of self-role accuracy, a measure of that pupil's general accuracy in evaluating how each of the others in his group evaluate him with reference to a given need-situation. A relatively small mean absolute Dg score indicates that he is generally accurate in his estimate as to how other group members appraise him with respect to his need satisfaction potential for a given social need.[2]

In an initial study (6), the reliability of average Dg scores for individuals was examined. The initial tryout, using a sample of 12 fraternity members tested twice the same evening (after an intervening, distracting task), yielded a test-retest correlation coefficient of .84. Forty-four percent of the 121 algebraic Dg measures for the 12 members were identical for the two administrations: 46 percent of them differed only one scale unit. A public school class of 22 sixth graders was tested twice with an intervening two-week period. The retest correlation coefficient was .66. Thirty-four percent of a random sample of 126 individual absolute Dg measures remained identical over the two-week period: 35 percent of them differed only one scale unit.

The reliability of the mean absolute Dg scores for pupils in the present New York State sample was estimated in terms of split-half coefficients of reliability computed for three classes of pupils, one from each of the three lowest grades studied. These coefficients were obtained by correlating the pupil's average accuracy score in predicting rating received from one half of their classmates with those received in predicting the other half and applying the Spearman-Brown formula for whole test reliability. The corrected reliability coefficients were .76, .76, and .88 for the fifth, sixth, and seventh grade classes respectively. These are coefficients based on individual mean accuracy scores. It may logically be assumed that the group means of individual mean accuracy scores employed in this study exhibit still greater stability.

The average absolute Dg measure is proposed as an index of the ratee's accuracy of percep-

tion of ratings he receives from other members of his group. As a dyadic measure, it incorporates an interaction of rater and ratee response behavior and is vulnerable to the biasing effects of centrality and variance of both estimates made and ratings received(5) In a major way, however, the proposed accuracy score differs from most prior social discrepancy measures. The Dg score is designed as a measure of the subjects' awareness of a rating he receives and incorporates no demands on either the rater or estimator to describe the actual "amount" of a trait the ratee possesses. The rater task is one of identifying, in terms of a paired comparisons procedure, each member of his group along his (the rater's) personal continuum of perceived need satisfaction potential. The ratee task is one of estimating the position at which his class mates "placed" him on that rating continuum. The ratee, as an estimator, is not required to consider his actual (as he might describe him self) need satisfaction potential nor the actual (felt but non-expressed) evaluation of his rater but only that rater's response. The accuracy of that rater's response does not enter here, only the accuracy of that ratee, as an estimator, in determining what that response was. This procedure is designed to measure how well group members are aware of how they are perceived (or rated) by others. An estimator's too generous description of the group's ratings of himself properly contributes to his Dg score. The relationship of the subject's accuracy score with his mean estimate or with the deviation of his mean estimate from his group's mean estimate is an aside here. It is more pertinent that the Dg measure, as an index of the subject's accuracy of social perception, be relatively independent of the mean of the ratings he receives and independent of the discrepancy of that rating mean from his group's rating mean. A lack of independence here would suggest that the proposed accuracy score is a function of how the group perceives or rates the estimator rather than how closely the ratee estimates each of the several ratings made of him.

The variance of the ratings an estimator receives describes the extent to which the group members do differentially rate him. In terms of the desired estimator-centered interpretations of the proposed accuracy score, differences in rating variance among raters should have minimal correlation with the Dg measure. The further question of the estimator's general tendency to estimate around his group's rating norm and thereby favorably promote his accuracy score may be examined in terms of correlation coefficients. A low relationship of the Dg score with differences between the pupil's mean estimate and the group's mean rating would describe an independence of the accuracy measure and the (inadvertent or planned?) successful approximation by the estimator of his groups' general level of rating.

An examination of the proposed mean absolute Dg score in terms of these foregoing considerations is presented in Table II. All correlations were computed separately by grade to vitiate extraneous developmental factors. The average r's (in the column at the right of the table) are based on weighted averages of grade correlations using r to z transformations.

The correlations between the mean of the estimates made by pupils and the mean of the rating received are reported in the first row of Table II. The generally low r's, average r of .08, suggest a definite lack of relationship between the two average pupil indices; the average estimate a pupil makes is relatively independent of the average rating his classmates make of him.

Examination of the within-grades relationships of the average rating received with the accuracy score (row 2, Table II), reveals a varying positive correlation of from .05 for the fifth grade to .52 for the eleventh grade. The average correlation over all grades is .27. Apparently, within some of the grades, there is a definite tendency for higher-rated pupils to be more inaccurate than pupils receiving lower ratings in estimating ratings made of them. However, in summarizing over grades the relatively low average r of .27 indicates a considerable independence of the Dg score with respect to the average rating received by the estimator.

The variance of the ratings which a pupil receives is similarly related to the accuracy measure, with an average r of .21 (row 3, Table II). Other conditions being equal, the pupil who receives a more varied set of ratings from his classmates will tend to earn a poorer accuracy score (higher Dg score).[3] Again, though, the extent of this relationship ($\bar{r}$ = .21) is far from compelling and suggests only a minimal interdependency of the Dg score and the variance of others' rating of him.

The within-grade correlations describing the degree of relationship of the pupil's accuracy score with the absolute deviation of the mean of that pupil's ratings from his groups' mean rating are presented in row 4 of Table V. These correlations relate to the problem of centrality: Are persons whose average ratings received most nearly approximate that of their group's average most accurate in estimating these ratings? More specifically, these correlations are a measure of the relationship of estimator dissimilarity as defined by his classmates' more extreme ratings of him, and estimator inaccuracy (the Dg score). With the exception of a fifth-grade correlation

coefficient of .46, these correlations tend to be reasonably small with an overall average r of .27. In general, only about seven percent of the pupil's variance on the accuracy measure may be accounted for by the non-centrality of his average rating received.

The last set of correlations describes the relationship of the accuracy score with the absolute difference between the pupil's general level of estimates and his group's general level of ratings. A high correlation here would support the contention that perception of, or attention to, or even accidental approximation of the more generalized rating level of the group, substantially promotes the individual's accuracy score. The r's in the last row varied from a high of .45 to a low of zero with an overall average r of .23 (row 5, Table II). In general, the accuracy score tends to be relatively independent or very slightly related to the pupil's departure as an estimator from the average ratings made by his group; on the average less than six percent of the pupil's variance on the accuracy measure may be accounted for by the extent of which his average estimate deviated from the central tendency of the rating made by his group.

In summary, the five sets of correlations described above contribute to a description of the Dg measure in terms of concomitant variations of the pupils' average estimate or accuracy score with indices based on the rating behaviors of others. The average estimate a pupil makes tends to be independent of the average rating he receives. There is a slight tendency for pupils receiving higher average ratings or receiving a more varied set of ratings to earn generally poorer accuracy scores; both the average of the ratings received and the variance of these ratings were found to have a low positive relationship with the derived accuracy score; average r's of .27 and .21, respectively. Similarly, a low positive relationship of the individual's accuracy score with the deviation of his ratee position from the group's general mean was found. The correlations of the difference between the pupil's mean estimate and the group's general level of rating with the pupil's accuracy score also revealed relatively low dependency or concomitance. The correlations reported in Table II are based on grade samples of 34 to 64 pupils per grade. There are no suggestions of grade trends for any of these correlations; in all cases, the between-grade variability of any set of correlations may be attributed to sampling error at the .05 level of confidence.

Postulated Relationships of the Perceptual Accuracy Variable

The thesis that accuracy of social perception, in this instance accuracy of self-role perception, is an acquired social ability and therefore responsive to the accumulated environmental experiences of the individual, leads to two critical factors determining this ability.

1. Experience in the social group: Other factors being equal, the greater number of contacts and/or amount of time spent in this and similar social groups, the greater the individual's accuracy of self-role perception.

2. General ability to abstract from, or normatively evaluate, social experience: Other factors being equal, the greater this ability, the greater the individual's accuracy of self-role perception.

The first factor, that of experience in the social group, incorporates the varying effects of motivation or drive state, varying both between individuals and within the same individual. The basket term, "opportunity for experience," insists not only on physical opportunity for attentiveness and/or participation, but also a non-apathetic approach to such opportunity. In a comparison of subgroups within a population, differences in this motivational factor are confounded within the subgroup delineation. The basic subgroup description considered in this study is that of amount of experience in social groups. The relatively homogeneous school setting permits operational restatement of this description in terms of grade placement.

A further refinement in subgroup definition is suggested by the recognition of differing "kinds" of experience. Within the class grade groupings, behavioral sex differences are strongly in evidence. Cultural stereotype sex roles are so much the law of the land that, for many, exception is aesthetically intolerable. The problem of whether or not these different experiences inaugurated by societal sex typing (and later by physiological persuasion) result in differing degrees of self awareness or self perceptions becomes a logical (and necessary) consideration.

In this study, the sex factor was considered in terms of two basic questions. First, do the sexes differ with respect to accuracy of self-role perception and do these differences maintain themselves for differing age groups? And second, do differences exist between estimates made of ratings received from same sex and opposite sex members?

TABLE III

PRELIMINARY DATA TO TESTING OF HYPOTHESES 1a AND 1b; DISTRIBUTION OF MEAN ABSOLUTE Dg SCORES BY SEX WITHIN GRADES

Grade	5	6	7	8	9	10	11	12
				Girls				
N	24	34	37	26	23	26	21	19
$\overline{X}$	2.12	1.93	1.98	1.50	1.40	1.48	1.20	1.35
σ	.427	.486	.624	.396	.357	.351	.314	.298
				Boys				
N	31	22	27	26	27	10	14	18
$\overline{X}$	2.22	2.02	1.65	1.46	1.42	1.34	1.10	1.58
σ	.563	.618	.559	.377	.507	.510	.264	.356
				Total				
N	55	56	64	52	50	36	35	37
$\overline{X}$	2.18	1.97	1.84	1.48	1.41	1.44	1.16	1.46
σ	.511	.543	.626	.387	.445	.406	.299	.347

TABLE IV

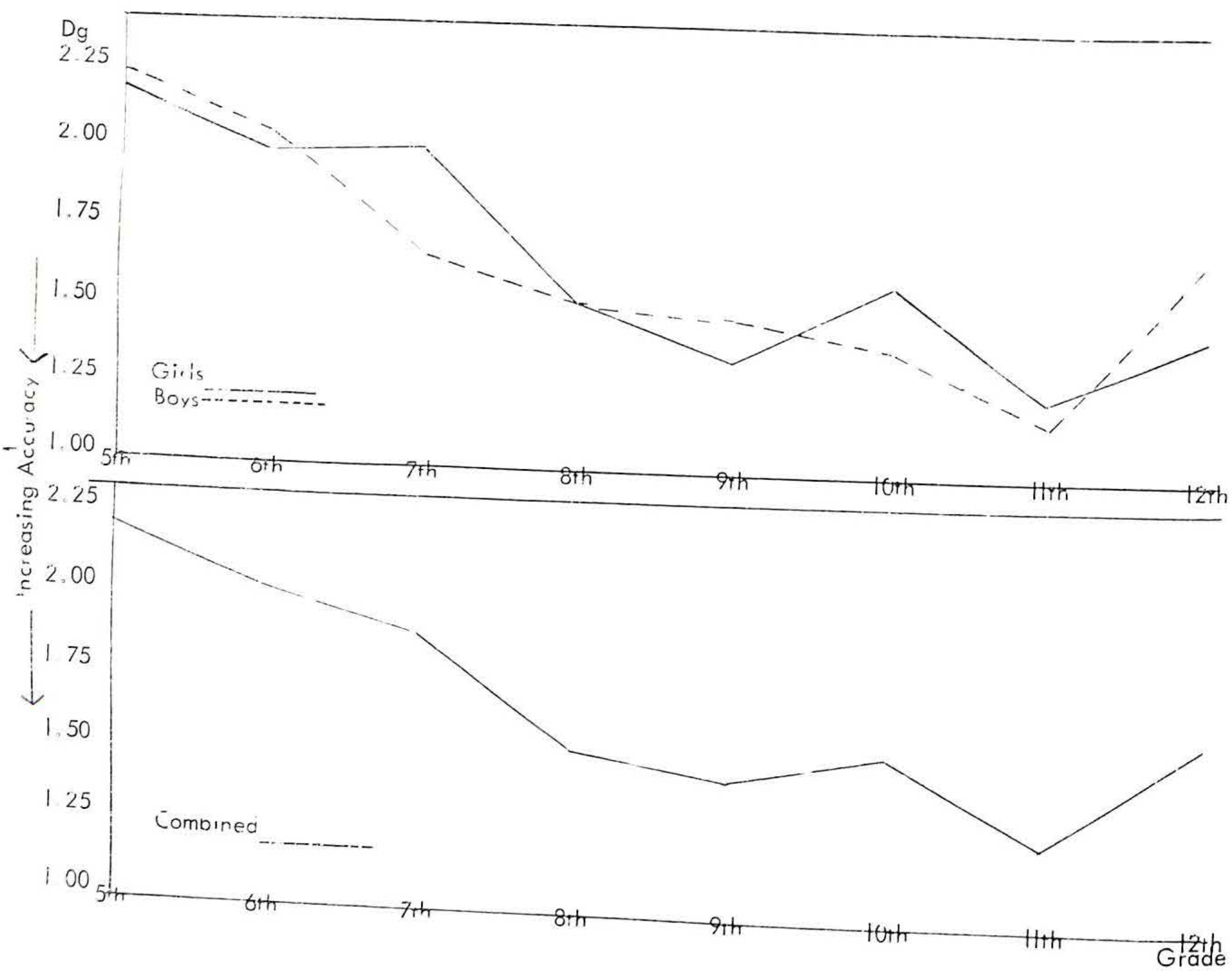
PRODUCT MOMENT CORRELATIONS BETWEEN MEAN ABSOLUTE Dg SCORE AND MENTAL AGE (AS DETERMINED FROM THE CALIFORNIA TESTS OF MENTAL MATURITY) FOR GRADES 5 - 12

Grade	5	6	7	8	9	10	11	12
N	55	56	64	52	50	36	35	37
$\sigma_{\lvert \overline{Dg} \rvert}$	.511	.543	.626	.387	.445	.406	.299	.347
σ_{MA}	15.4	15.2	18.0	18.5	11.5	12.1	17.2	15.8
r	-.35*	-.03	.06	.11	-.03	-.14	-.15	.08

*Significant at the .05 level

FIGURE 1

GRADE MEANS BASED ON MEAN ABSOLUTE Dg SCORES FOR BOYS AND GIRLS SEPARATELY AND COMBINED



The second postulated factor, general ability to abstract from social experience was considered in terms of mental age score. The following six hypotheses pertinent to the foregoing discussion, and related research (1, 2, 7, 13), may now be explicitly stated:

H:1a. With increasing grade, the mean absolute Dg score will decrease.

1b. There will be a difference in decreasing trends for the two sexes.

H:2a. Within each grade, there will be a negative relationship between MA and the mean absolute Dg score.

2b. These negative relationships will become less pronounced with increasing grade.

H:3a. Within each grade, the mean absolute Dg score based upon same sex ratings will be lower than those based upon opposite sex ratings.

3b. These differences will decrease with increasing grade.

Results

The preliminary data to the testing of the first set of hypotheses (1a, b) are presented in Table III. Hypothesis 1a was examined in terms of an analysis of variance approach. Snedecor's approximate test of mean differences under conditions where the assumption of homogeneity cannot be met (16, pp. 287-89) yielded an F' of 28.10 significant at the one percent level for seven and 153 degrees of freedom. The product moment correlation coefficient between grade and grade mean absolute Dg score was -.50. The hypothesis of increasing accuracy of self-role perception with increasing grade was supported.

The departure of the trend of decreasing means from linearity was examined following McNemar's (14) presentation involving η^2 and tests of curvilinearity. The $\eta^2 = .3089$ ($\eta = .556$). The Pearson product moment correlation coefficient squared is .2460 ($r = -.496$). Both the η and r are significantly different from zero at the one percent level of significance. The F ratio for testing the significance of the difference between η^2 and r^2 was 5.72 with six and 377 degrees of freedom, significant at the one percent level. The relationship of decreasing accuracy with increasing grade may be described as non-linear. The actual extent of this non-linearity may be seen in Figure 1.

The difference in decreasing trends of mean absolute Dg scores with increasing grade for the two sexes was examined by means of a factorial design. In view of the disproportionality of cell frequencies, Snedecor's method of fitting constants was followed in correcting the sums of squares for grades and sex (16, pp. 388-91).

The F ratio for testing the significance of the difference between means for the two sexes was less than one. The F ratio for testing the significance of a grade-sex interaction failed to meet the five percent level of significance ($F = 1.64$ with seven and 369 degrees of freedom, $p. \doteq .10$). The difference between the means for the boys and girls over all grades and the differences observed at the various grade levels may be attributed to chance patterning. The trends for the two sexes may not be described as statistically different either in terms of vertical displacement or slope. Boys and girls are not differentiated in terms of the experimental (Dg) measure of accuracy of self-role perception (see Figure 1).

With the singular exception of a correlation coefficient of -.35 in the fifth grade, no support was found for the hypothesized negative relationships of mental age and the mean absolute Dg score for the various grades. None of the product moment correlation coefficients for the upper seven grades were significant at the five percent level. The largest was a coefficient of -.15 for a sample of 35. In view of this preponderance of exceedingly low r's, hypothesis 2b was not considered further. The eight correlations are reported in Table IV.

The last set of hypotheses involve differences between the pupil's accuracy in estimating the ratings he received from same sex members and his accuracy in estimating the ratings he received from opposite sex members. Differences between the mean absolute Dg scores based on same and opposite sex members were recorded for each pupil and summed for each grade. The significance of the mean of these differences for each grade was tested, using a t-test for matched pairs. The results of these tests are presented in the lower third of Table V.

It is obvious that, with one minor exception at sixth grade, the total grade differences in accuracy of estimating ratings received from same and opposite sex members is in the opposite direction to that hypothesized. Three of these differences are significant at the one percent level. However, no trend in these differences from lower to higher grades is discernible. η^2 was computed as .046, r as zero.

To explore further the phenomena of differences in accuracy of perception of girls' and of boys' ratings of oneself, the data were analyzed separately for each sex. The upper two thirds of Table V summarizes this analysis. A similar pattern of significance/non-significance as that for analysis of the combined groups is evident.

The mean accuracy scores for each grade in terms of perceptions of same sex and opposite

sex ratings of self are plotted in Figure 2 for girls and boys separately and combined. The greater consistency of the data for the boys is at once apparent, as are differences between the two sexes at the various grade levels.

Discussion

The hypothesis that accuracy of self-role perception is positively related to grade placement is well supported by the data in this study. This overall finding agrees with that of Ausubel, Schiff and Gasser, who found relatively consistent increasing trends in sociempathetic ability over an age of eight to 17 years (2). Dymond, Hughes and Raabe had earlier reported greater perceptive ability in older children in their study of seven and eleven-year-olds (7). Lundberg and Dickson found similar trends in accuracy in predicting sociometric reciprocation in their high school study (13). Argument that the Dg trend might be due to the increasing mental age level of the pupil is considerably weakened by the failure to find continuing relationships of mental age with the Dg measures within each grade.

It should be added that the equivalence of age with grade placement would be procedurally unsound for this report. The major hypothesis of an increasing accuracy with increasing grade placement was developed in terms of the postulated relationship of experience in the social group and accuracy of self-role perception. The approximate three-year range of chronological age within each grade level permitted a tentative examination of the relationship of accuracy and age experimentally controlling for grade placement. With the exception of an r of .40 and of .26 for the fifth and sixth grades respectively, the correlations of the mean absolute Dg score and the pupil's chronological age in months were all zero or near zero. Further inspection of the older pupils in these two lowest grades tested identifies them as a well below-average subgroup with respect to I.Q. as well as Dg score. Adequate explanation of this description of the greater inaccuracy of the older, low I.Q. child in the lower grades (five and six) requires further examination of the interpersonal relations and of the rating response patterning of these children. In general, it may be stated that except for older pupils in the lower grades tested, zero relationships of social accuracy with age, controlling for grade placement, were found.

The grade trend in mean absolute Dg scores for combined sexes is graphed in the lower part of Figure 1. The noticeable reversal toward poorer social perception, i.e., higher mean absolute Dg scores after grade eleven, needs explanation. Lundberg and Dickson's report describes a similar break in pattern. Their data were based on reciprocations of nominations of best friends. Since their students were all able to choose friends from the entire span of high school students, a contributing explanation might be suggested in terms of the argument that seniors "know" fewer others since they would be in a smaller class and that they have no upper classmates to choose from. It may well be that fewer of their best friends are in the high school student group as compared with friends of students in other grade groups.

An offered explanation of the present reversal in class mean accuracy scores follows a somewhat similar vein. It is not unreasonable to assume that the seniors, when compared to their younger schoolmates, have a considerably wider outside reference population, have more varied interests, experiences, acquaintances, and perhaps, more pressing social contacts in terms of marriageable mates, potential or planned college, army, or business chums, and the like. This renders their judgmental frame more difficult to appraise. The seniors' behavior with their class members may be more forced than genuine with many "real" friendships or comradeliness now reserved for community chums no longer in school. The seniors' behavior may be more socially polished, evincing friendships which are less detectible in terms of their actual strengths than are those of younger pupils.

This is, of course, tentative explanation. The fact that Lundberg and Dickson found a similar pattern suggests a real, non-methodological reversal. Further inspection of sex differences (upper part, Figure 1) points to a greater increase in inaccuracy for boys. Figure 2, (middle part), suggests that much of this greater inaccuracy for boys is due to a sharp decrease in accuracy of predicting how their own sex members rate them. To some extent, this receives support from Ausubel and Schiff's eleventh grade data (1) indicating that the sociometric attitudes held by eleventh grade girls are more apparent than those held by eleventh grade boys. But sex subgroups aside, the increased inaccuracy, or larger mean absolute Dg score of the twelfth graders in this current study is definitely not supported by the Ausubel, Schiff and Gasser developmental report (2) which found that the greatest gain in perception of own status occurs between the eleventh and twelfth grades. Major methodological differences in terms of selected rating format, the rating and continuum, and between the mean absolute Dg score and Ausubel's group correlation coefficient vitiate a more formal comparison of these conflicting results.

The major increase in accuracy at about the seventh and eighth grades is not unexpected in terms of continuing research reports describing

FIGURE 2

GRADE MEANS OF MEAN ABSOLUTE Dg SCORES BASED ON ESTIMATES OF RATINGS RECEIVED FROM SAME SEX AND OPPOSITE SEX MEMBERS FOR BOYS AND GIRLS, SEPARATELY AND COMBINED

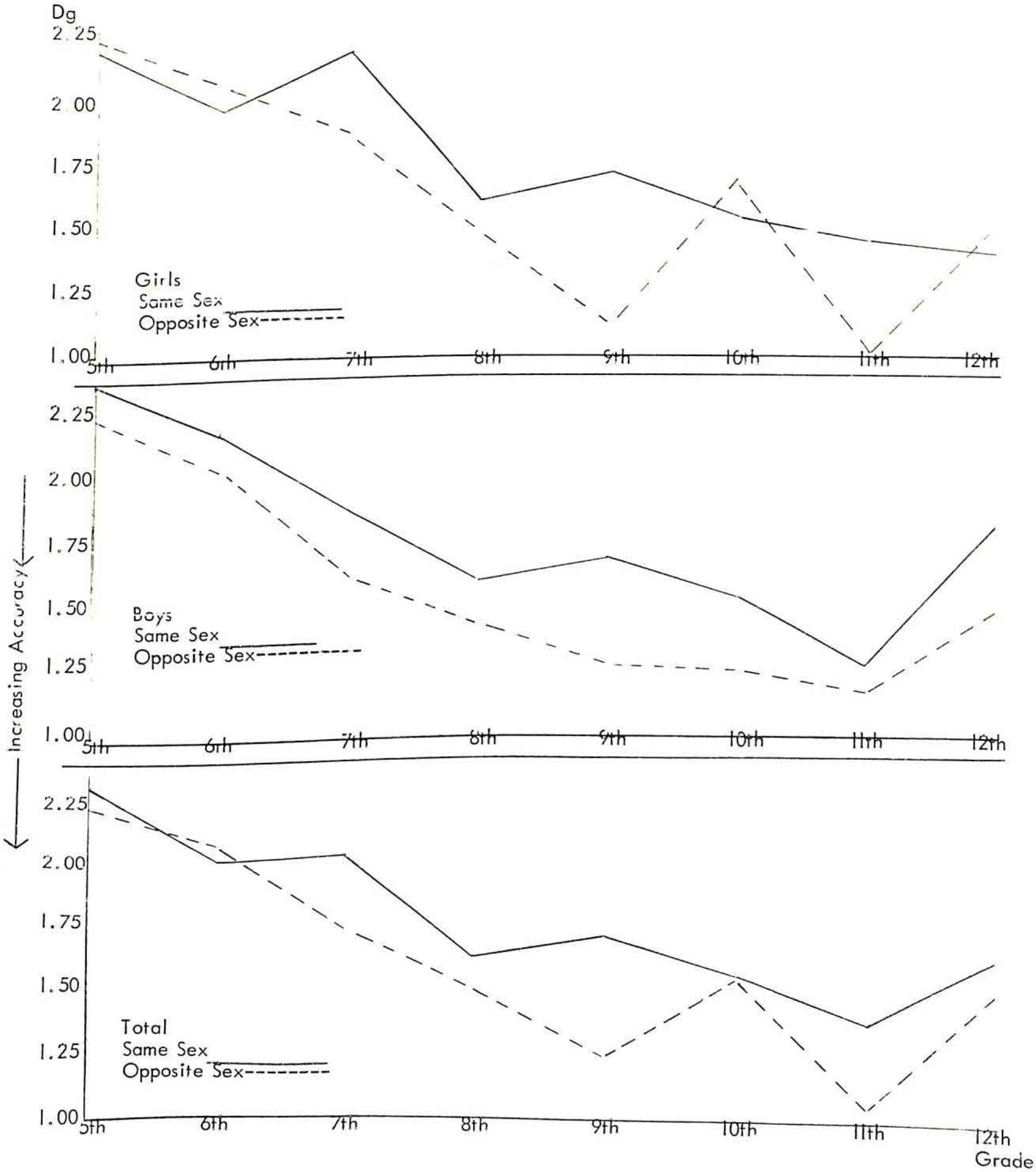


TABLE VI

MEANS OF ESTIMATES, RATING, AND ABSOLUTE Dg SCORES FOR GRADES 5-12

Grade	5	6	7	8	9	10	11	12
Estimates Made	3.46	3.36	2.74	2.73	2.58	2.52	2.49	2.70
Ratings Received	3.52	3.43	3.04	2.86	2.82	2.78	2.54	2.87
Absolute Dg Scores	2.18	1.97	1.84	1.48	1.41	1.44	1.16	1.46

TABLE V

THE RESULTS OF TESTS OF SIGNIFICANCE OF DIFFERENCES BETWEEN MEAN ABSOLUTE ACCURACY SCORE $|\bar{D}g|$ BASED ON ESTIMATES OF RATINGS RECEIVED FROM SAME AND OPPOSITE SEX MEMBERS

(Negative Mean Differences indicate greater mean accuracy with respect to opposite sex members than with respect to same sex members.)

Grade	5	6	7	8	9	10	11	12
				Girls				
N	24	34	37	26	23	26	21	19
$\bar{D}$	.061	.112	-.289	-.161	-.547	.097	-.429	.077
σ_D	.876	.633	.657	.417	.643	.868	.502	.752
t	.33	1.02	2.64*	1.93	3.99*	.56	3.82*	.43
				Boys				
N	31	22	27	26	27	10	14	18
$\bar{D}$	-.104	-.100	-.291	-.147	-.433	-.313	-.106	-.326
σ_D	1.124	.633	.704	.737	.510	.584	.486	.733
t	.52	.72	2.11*	1.00	4.33*	1.6	.78	1.83
				Total				
N	55	56	64	52	50	36	35	37
$\bar{D}$	-.032	.029	-.290	-.154	-.485	-.172	-.300	-.119
σ_D	1.027	.642	.677	.598	.578	.820	.520	.770
t	.23	.34	3.40**	1.84	5.88**	.12	3.36**	.93

* Significant at the .05 level ** Significant at the .01 level

adolescent behavior. Kuhlen and Lee, in particular, emphasize "emerging heterosexual interests from the sixth grade on." (12) In describing increased heterosexual relations, they further note that most of this increase is between grades six and nine.

The displacement in sex trends at these grade levels is more difficult to explain. The data clearly show (see Figure 1, upper part) that the boys, as a group, have an earlier "spurt" in increased accuracy than girls. The major increase in accuracy for the boys is a grade earlier than for girls, beginning between the sixth and seventh grades; the girls catch up to them in the eighth grade. It has been expected that the earlier pubescence of the girls (the Harvard Growth Study (15) suggests that the average age of pubescence for girls is about thirteen to thirteen and a half years; boys about two years later) and their accompanying older social age would be reflected in terms of greater social self-awareness, i. e., lower mean absolute Dg scores. This was not the case; rather the opposite trend was noted.

Partial interpretation may be found from examination of the two pairs of curves based on accuracy of estimates of ratings received from same and opposite sex members (Upper two thirds, Figure 2). The major portion of the reversal in mean absolute Dg scores for the girls is evidently due to the decreased accuracy of the seventh grade girls in estimating ratings received from their own sex members. The general trend of means for all grades of girls (excepting the tenth grade, see below) descriving their increased accuracy with respect to opposite sex members is very similar to that for all grades of boys.

The observed increase in inaccuracy for seventh grade girls in estimating ratings received by same sex classmates suggests an intensified process of reappraisal or redefinition of interperson roles for this group. Any judgmental shifting by his peers along the need satisfaction potential scale (in this instance, need-succorance) requires a corresponding reappraisal by the ratee or estimator to maintain an equivalent accuracy score. Unstable or recent shifting by the rater promotes inaccuracy on the part of the ratee. The seventh grade girls in this sample are probably well divided between pre and post pubescence; the grade mean age is thirteen years with a range of from twelve to fifteen years. The fact that no similar sharp reversal is evident for any age group of boys suggests a more stabilized or immediately apparent peer evaluation for the boys with respect to the selected rating continuum.

A second sharp distortion in Figure 2 (upper third), that of the very marked increase in inaccuracy of tenth grade girls in predicting boys' ratings of themselves, tentatively resolves itself in terms of an unexpected sampling limitation. In the tenth grade, there were only eleven boys enrolled in the grade with 27 girls; the girls were definitely at a disadvantage socially. After randomly forming two testing sections, not counting two later absentees, one section contained six boys with thirteen girls and the second section, four boys and thirteen girls. Particularly in the latter section, one "unreciprocating" male rater, especially the extreme "error" of an unrewarded crush, could seriously skew the opposite -same sex data for the group. In any case, the somewhat extreme boy-girl ratio and the small N discourages generalization.

In general, Figure 2 depicts both sexes favoring opposite sex members in terms of awareness of how they are rated, though the data for the girls is much less consistent than that of the boys.

The fact that no significant relationship was found between the mental age score and accuracy of self-role perception for any grade level except that of the fifth grade fails to support the hypothesized negative relationship of these two variables. The fallibility of mental age scores derived from self-administered group tests contributes in part to these low r's both in terms of attendant errors of measurement and increased stress on academic reading abilities. However, a more substantial argument is that a sufficiency of experience and intellectual level is possessed by the majority of the pupil sample tested and that measures of a general intelligence factor beyond that base line no longer serve to distinguish the socially aware or accurate from the socially inaccurate. The higher correlation coefficient obtained for the lowest grade tested in this study endorses this interpretation. Further support for this "sufficiency" argument would require a more extensive testing of pupils in the lower mental age ranges as well as pupils in the lower grades. Such a testing program is anticipated by the senior author.

A final consideration relating to the concomitant variations of the basic estimating and rating indices is warranted here. In reporting a trend of a discrepancy measure; it is pertinent also to examine the trends of the parent measures, in this instance, the pupil estimates and ratings. The grade means of the pupil estimates and ratings and of their absolute Dg scores are reported in Table VI. It is immediately apparent that grade means of both estimates and of ratings follow a negative trend with increasing grade placement very much like that of the accuracy measure. The product moment correlations for these indices with grade placement, using an eight-point scale for the latter, of -.32, -.32 and -.50, respectively, provide a comparison of the degree of

linear trends for these three indices.

The fact that pupils in the higher grades tend to use more of the lower end of the scale for both their ratings and their estimates suggests the possibility of a contributing artifactual relationship of the derived accuracy measure and grade placement.[4] Groups with a general tendency to give lower ratings may be more apparent in their rating behavior than are groups with a general tendency to rate higher independent of grade placement. Aside from possible actual "apparency," the reduced variance associated with ratings or estimates nearer the terminal end of a rating scale would in itself suggest smaller discrepancy (Dg) scores.

The trend of the accuracy measure with grade was further examined in terms of partial correlation coefficients. The unadjusted correlation between grade and accuracy (see above) is -.50. The correlations of the accuracy measure with the average estimate made and with the average rating received for all grades combined are .54 and .38 respectively. Holding constant the effect of average estimates made the relationship of grade and accuracy reduced to -.41, a change of .09 correlational points. Holding constant the effect of average ratings received the relationship reduced to -.43, change of .07 correlational points.

In general, it may be concluded that the concomitant grade trends of the estimating and rating responses do not compel the obtained grade trend for the accuracy measure. Though the accuracy score is moderately correlated with pupil level of estimating, and to a lesser degree, with the level of the ratings a pupil receives, these relationships statistically account for only a small portion of the observed grade trend in accuracy.

Summary

One aspect of social perception, namely, that of accuracy self-role perception, was studied in terms of a public school sample of 385 pupils in grades five through twelve. The accuracy variable was operationally defined as a distance function describing the difference between a rating received from another and expectation (perception) of that rating. The Syracuse Scales were employed as a basic format for eliciting the initial rating, described in terms of a rater's perception of the need satisfaction potential of the ratee with respect to a given social need. Social need "succorance" was chosen as a social need least biased for this study. A mean absolute discrepancy, or Dg score, was obtained for each pupil.

The pupil's general level of estimating was found to be independent of the general level of ratings he received. The relationships of his proposed Dg score with several rating indices were examined and were found to be moderately and negatively related to his accuracy in estimating ratings he receives.

Grade means were examined and the hypothesis of increased accuracy of self-role perception with increasing grade was supported. A non-linear trend was reported with a reversal (increasing inaccuracy) evident after the eleventh grade. No overall sex differences in either mean accuracy or in trend with grade were found.

Both sexes were found to be more accurate in perceiving ratings made of them by opposite sex members than by same sex members, though girls were less consistent in this respect than were boys. With the exception of the fifth grade, no significant relationship was found between mental age (California Tests of Mental Maturity) and the accuracy score at any grade level. The problem of possible artifact in the reported grade differences in the accuracy score related to concomitant variations in grade levels of estimates and of rating was considered.

FOOTNOTES

1. For review and discussion of these varying approaches, see Borgatta (3), Bronfenbrenner, Harding, and Gallwey (4), Gage (8), Taft (17), and Tagiuri and Petrullo (19).
2. It is assumed that the invalidating factors of pupil reluctance and non-cooperation, tendencies on the part of the ratee to deliberately modify or deny certain awarenesses incompatible with his preferred self-concept and security have been reasonably minimized by the confidential nature of the experimental testing situations. All participants were told that their responses would be kept anonymous except for the required statistical handling.
3. This relationship is in accord with Bronfenbrenner, Harding and Gallwey's discussion of the component perceptual skills of "interpersonal sensitivity" and "sensitivity to the generalized other." (4) It is reasonable to expect that the estimator's insensitivity to individual rater differences is a less crucial component of his accuracy score in estimating a more homogeneous set of ratings and that this same "insensitivity" will result in greater inaccuracy with increased rater variance.
4. As may be evident to the reader from earlier tables, both the number of pupils per grade and the size of the inter-rating subgroups

generally decrease with increased grade placement. The effect of increased length of the rating-estimating task for the larger groups was examined by comparing Dg scores obtained at the beginning and at the end of the testing period. Only minor chance differences were found here.

The problem of grade size as a determinate of accuracy in the sense that pupils in less populous grades have greater opportunity for contact with all their classmates with a resulting greater awareness of other's ratings of themselves was minimized in this study by sampling within a small single-building school affording a maximum of sustained pupil interactions. (See Subjects and Procedure).

REFERENCES

1. Ausubel, David P. and Schiff, H. "Some Intrapersonal and Extrapersonal Determinates of Individual Differences in Sociempathy Ability Among Adolescents," Journal of Social Psychology, XLI (1955), pp. 39-56.
2. Ausubel, David P., Schiff, H., and Gasser, E. "A Preliminary Study of Developmental Trends in Sociempathy: Accuracy of Perception of Own and Other's Sociometric Status," Child Development, XXIII (1952), pp. 111-28.
3. Borgatta, Edgar F. "Analysis of Social Interaction and Sociometric Perception," Sociometry, XXVIII (1954), pp. 7-32.
4. Bronfenbrenner, Urie; Harding, John, and Gallwey, Mary. "Measurement of Skill in Social Perception," in McClelland, David C.; Baldwin, Alfred L.; Bronfenbrenner, Urie, and Strodtbeck, Fred L. Talent and Society. Princeton, New Jersey: D. Van Nostrand Co., Inc., 1958.
5. Cronbach, Lee J. "Processes Affecting Scores on 'Understanding of Others' and 'Assumed Similarity'," Psychological Bulletin, LII (1955), pp. 177-93.
6. deJung, John E. "The Measurement of Accuracy of Self-Role Perception," unpublished doctoral dissertation, Syracuse University (1957).
7. Dymond, R.; Hughes, A. S. and Raabe, V. L. "Measurable Change in Empathy with Age," Journal of Consulting Psychology, XVI (1952), pp. 202-6.
8. Gage, N. L. "Explorations in the Understanding of Others," Educational and Psychological Measurement, XIII (1953), pp. 14-26.
9. Gardner, E. F. and Thompson, G. G. Group Administered Forms of the Syracuse Scales of Social Relations. Annual Report of Research Supported by the Office of Naval Research, Department of the Navy, (mimeographed paper), Syracuse, New York, Institute of Research, Syracuse University, 1957.
10. Gardner, E. F. and Thompson, G. G. Social Relations and Morale in Small Groups, New York: Appleton-Century-Crofts, Inc. 1956.
11. Gardner, E. F. and Thompson, G. G. Syracuse Scales of Social Relations, Yonkers, New York: World Book Co., 1959.
12. Kuhlen, R. G. and Lee, B. J. "Personality Characteristics and Social Acceptability in Adolescence," Journal of Educational Psychology, XXXIV (1943), pp. 321-40.
13. Lundberg, G. A. and Dickson, Leonora. "Inter-ethnic Relations in a High School Population," American Journal of Sociology, LVIII (1952) pp. 1-10.
14. McNemar, Quinn. Psychological Statistics (2nd Ed.), New York: John Wiley and Sons, Inc., 1955, pp. 262-75.
15. Shuttleworth, Frank K. "The Physical and Mental Growth Curves of Girls and Boys, Age Six to Sixteen, In Relation to Age of Maximum Growth," Monographs for the Society of Research in Child Development, IV (1939), p. 291.
16. Snedecor, G. S. Statistical Methods (5th Ed.) Ames, Iowa: Iowa State Press, 1956.
17. Taft, Ronald. "The Ability to Judge People," Psychological Bulletin, XLII (1955), pp. 1-23.
18. Tagiuri, R.; Blake, R. R. and Bruner, J. S. "Some Determinants of the Perception of Positive and Negative Feelings in Others," Journal of Abnormal and Social Psychology, XLVIII (1953), pp. 585-92.
19. Tagiuri, R., and Petrullo, L., Eds. Person Perception and Interpersonal Behavior, Stanford, California: Stanford University Press, 1958.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 1, September 1962)

AN EXPERIMENTAL STUDY OF TRAFFIC SAFETY FILMS

H. H. McASHAN**

Titusville, Florida

EACH YEAR there are about 40,000 deaths from automobile accidents and several hundred thousand people are seriously injured. This traffic toll is greater than that of several of our most serious diseases. Despite recognition by various groups of our safety needs, and of the seriousness of our traffic problem, the accident rate has continued to increase each year. Over 300 films have been produced on traffic safety. These films are being shown to a large number of viewers each month with little evidence that traffic safety is being improved.

Film research studies have shown that the dramatic type of film most often produced actually may be very ineffective in changing a person's attitudes and habits. Thus, a careful analysis of the learning from the content of traffic films could have an important bearing on determining how effective a traffic safety film is upon the shifting of attitudes and the changing of driving performance. The purpose of this study is to make factual, attitudinal, and behavioral analysis of traffic safety films. A second objective is to investigate the factors involved in determining driver behavior.

Experimental Design

The groups for the study were drawn from a population of sophomore, junior, and senior students at Michigan State University. The samples used consisted of 165 students from classes representing the departments of Education, Health, Physical Education, Audio-Visual, Police Administration, and Highway Traffic Safety. The selection of these people was by both the purposive and random assignment methods. The students were divided into four groups by random assignment methods which indicated into which of two control or two experimental groups he, or she, was placed.

In order to secure information on the effects of a traffic safety film upon the target audience, an instructional and demonstrational film "Driving At Night" was selected for use in the experiment after the previewing of thirty traffic safety films. A matched-group technique was employed as the best method of determining what changes among the sample population could be attributed solely to the effects of the experimental film. This technique was carefully prepared with a view to eliminating bias in its results and confusion in the interpretation.

Tests and measures were developed to evaluate the audience reaction to knowledge, attitudes, and driving skills on both an immediate and retention basis. These tests were administered to the sample subjects in such a manner that no student was required to take any test more than one time. The test of knowledge was coded to include the major informational points in the film. The semantic differential attitude scale used was based upon seven concepts thought to be the most significant contributions obtainable from this film. Charles Osgood and others have made valuable contributions toward the development of this type of test for the measurement of attitudes (2, p.192). A series of nine skill items were included in the test of driver skill or performance. These skill items were carefully chosen as being representative of the types of skill which the film might be expected to improve. Three additional test items were included in this test in an effort to measure the population's driving ability prior to the experiment.

Before the beginning of the experiment, the writer completed a behavioral analysis of the film "Driving At Night." This "behavioral an-

---

* The research reported herein was supported by a grant from the U.S. Office of Education, Department of Health, Education and Welfare.

** Director of Research and Projects.

alysis'' technique is a new device used to predict what effects a film may have upon an audience. It was developed from information which was secured by personal interviews of one hour's duration and the direct observation of driver behavior through street corner observation. The findings derived from these activities indicated that there are various factors which contribute toward driving behavior. Therefore, these factors must be influenced or changed if a person's driving behavior is to be improved. The development of the behavioral analysis technique takes these factors into account and suggests a new method for evaluating the effect of an audio-visual message upon the behavior of the viewers. The experimental design may be viewed graphically by reference to Figure 1.

The method or procedure of administering the experimental design was as follows:

1. Administer the "Behavioral Analysis" predictive instrument to five audio-visual and five traffic safety experts before the beginning of the experiment.
2. Divide the sample population into two control and two experimental groups by random sample procedures.
3. Show Experimental Groups III and IV the film "Driving At Night." At the same time, Control Group I and Control Group II were viewing a non-related film entitled "The ABC of Internal Combustion Engines."
4. Immediately following the showing of the above-mentioned films, Control Group I and Experimental Group III were administered tests concerning attitudes, achievement, and driving skills.
5. Six weeks after the showing of the films, Control Group II and Experimental Group IV were given tests on attitudes, achievement, and driving skills.
6. Analysis of the data was made following the testing of Control Group II and Experimental Group IV.
7. At the completion of the experiment, the results obtained from the sample population were compared with the predictions of each of the groups of experts representing the audio-visual and traffic safety fields.

Test Results and Analysis of Data

Complete data was obtained on 165 students. Of this group, 57 were in Control Group I which saw the non-related film; 28 were in Control Group II which saw the non-related film; 54 were in Experimental Group III which viewed the experimental film, and 26 were in Experimental Group IV which viewed the experimental film.

In order to provide a sound basis for inference, the hypotheses in this study were formulated in exact terms. The first two hypotheses in this study stated that there is no immediate or permanent difference in the factual knowledge of experienced drivers who have seen the film "Driving At Night" and of experienced drivers who have not viewed this film. The hypotheses seek to determine if the mean proportion of correct answers for individuals in Experimental Groups III and I are greater than the mean proportion of correct answers for individuals in Control Groups I and II immediately after and six weeks after viewing the films.

Since all groups used the same test, the individual scores made by the students on each test were used in the statistical analysis. First, the F test was used to determine if the variances of the matched populations might be presumed equal. Second, the null hypotheses were tested by the t-test of significance. The result of the F test revealed that there was no statistically significant difference between variances of the two groups of matched populations. The probability was between F.025 and F.975, that is, the chances of the variances of the matched groups being different were less than five in 100. Thus, the t-test of significance of the differences between the means of the control and experimental groups was made assuming that differences among the respondents were negligible.

The t-test revealed that the difference between the means of Experimental Group III and Control Group I was significant with a probability of 99.9 in 100. The difference between Experimental Group IV and Control Group II was significant with a probability of 95 in 100. This indicated that both experimental film groups and non-related film groups were significantly different in their learning and in their retention of factual knowledge as a result of viewing the films. On the basis of these results, therefore, the writer rejected the first and second null hypotheses. Table I lists the results of the analysis of the data concerning learning.

The third and fourth null hypotheses state that there is no immediate or permanent difference in the driving attitudes of experienced drivers who have seen the film "Driving At Night," and of experienced drivers who have not viewed this film. These hypotheses were to determine if the mean score for individuals in Experimental Groups III and IV were different from the mean score for individuals in Control Groups I and II immediately and six weeks after viewing the experimental and non-related films. The semantic differential attitude scale was based upon seven concepts thought to be the most significant contribution obtainable

FIGURE 1.

EXPERIMENTAL DESIGN

Control Group I	Control Group II	Experimental Group III	Experimental Group IV
1. Show non-related film	1. Show non-related film	1. Show experimental film	1. Show experimental film
Test immediately	Test six weeks later	Test immediately	Test six weeks later
1. Attitude	1. Attitude	1. Attitude	1. Attitude
2. Achievement	2. Achievement	2. Achievement	2. Achievement
3. Skills	3. Skills	3. Skills	3. Skills

TABLE I

RESULTS OF F TESTS, t-TESTS, AND MEAN SCORES

Item:	Matched Groups I and III		Matched Groups II and IV	
F Test:	$P < .05$		$P < .05$	
Mean Scores:	(I) 21.877	(III) 28.963	(II) 23.429	(IV) 26.154
t-Test:	$P > .999$		$P > .95$	

from the film. These concepts were: (a) Night Driving Skill, (b) Headlight Limitations, (c) Night Driving Visibility, (d) Night Driving Fatigue, (e) Proper Headlight Use, (f) Pedestrians, and (g) Driving at Night. There were four evalation scales used for each concept which were: good - bad, safe-dangerous, nice-awful, and wise-foolish. The F test of significance for each concept was used, as was the case in the tests for knowledge.

The results of the F tests revealed that there was no statistically significant difference between variances of the matched populations. Thus, the t-tests of significance of the difference between the means of the control and experimental groups were made by assuming that the differences between the respondents were negligible. The t-tests of significance between the mean scores of Control Group I and experimental Group III were found to be not significant within a probability of 95 in 100 for any of the concepts. The same results were obtained on the t-tests between the mean scores of Control Group II and Experimental Group III on all but the concepts of "night driving skill" and "night driving fatigue." These two concepts revealed that there is a positive shift in attitude between the experimental and control groups.

The analysis of the results of the tests for both immediate and permanent attitude shift may be interpreted as being meaningful in either of two directions. First, did any significant attitude shirt occur? Second, did the results obtained by the whole of the concepts allow us to accept statistically or to reject the null hypotheses stated?

The former case may be answered in the affirmative. While there was no shift in attitude detected immediately after viewing the film, a positive shift was established for the concepts "night driving skill" and "night driving fatigue" by the analysis concerning the permanent effects of the film upon attitudes. This attitude shift represents the same type of "sleeper effect" which was found by Hovland, Lumsdaine, and Sheffield in their study of the motivational film "The Battle of Britain" (3, pp. 182-200). This represents a significant finding since the experimental film utilized approximately fifty per cent of its footage in visualizing the importance of skill and the awareness of factors involving fatigue. Apparently the concepts representing hazards of visibility and general problems of driving at night were not affected by this film presentation. The differences found were evidently not due to preferential treatment by the producers of the film, since the presentation of all the concepts was by both visual and verbal instruction of about equal quality.

The second question is summed up in the following statistical analysis: The F test of the total concepts revealed no statistically significant difference between the variances of any of the matched populations. The t-test revealed that the difference between the means of the matched experimental and control groups was not significant within a probability of 95 in 100. Therefore, on the basis of all of the evidence, the writer accepts the null hypotheses. Table II lists the results of the analyses of the attitude data.

The fifth and sixth null hypotheses stated that there is no immediate or permanent difference in the measurable change in driving performance of experienced drivers who have seen the film "Driving At Night" and experienced drivers who have not viewed this film. The objective of these hypotheses is to determine if the mean proportion of correct answers for individuals in Experimental Groups III and IV are greater than the mean proportion of correct answers for individuals in Control Groups I and II immediately and six weeks after viewing the experimental and non-related films. More specifically, the writer attempted: 1) to compare the two control groups as to nine item scores; 2) to ascertain if the more skilled drivers performed better than the less skilled drivers on the nine-item test, regardless of the film viewed or the time the driving test was taken, and 3) to determine if Experimental Groups III and IV did better than Control Groups I and II on the nine-item test.

As was the case in the preceding hypotheses, the F test and t-test were again used for their statistical effects. The F test revealed that there was no statistically significant difference between the variances of any of the groups investigated. The differences between the means of each matched group were not significant within a probability of 95 in 100. These test results indicate that the experimental film groups and the non-related film groups had no significantly different change in their driving skill either immediately or six weeks after viewing the films. Therefore, on the basis of the evidence derived, the writer accepted the null hypotheses. Table III lists the results of the analysis of the data concering driving performance.

Results and Analysis of Tests

The seventh hypothesis stated that it is possible to predict, by the use of a behavioral analysis, the measurable change in knowledge, driving attitudes, and driving skills which will occur as a result of viewing a traffic safety film. This hypothesis sought to determine how the estimates of the panel of experts correlated with the actual performance of students on tests of learning, attitude, and potential driving behavior. In the analysis of this hypothesis, the writer treated each compo-

TABLE II

THE RESULTS OF THE F TESTS, t-TESTS, MEAN SCORES, AND HYPOTHESES DISPOSITIONS FOR ATTITUDE SHIFTS

Concept	F Test	Mean Scores I	Mean Scores III	t-Test	Hypothesis Disposition
Immediate Attitude Shifts					
1. Night Driving Skill	P< .05	3.82995	4.05500	P > .95	Accept
2. Headlight Limitations	P< .05	4.45614	4.45370	P > .95	Accept
3. Night Driving Visibility	P< .05	3.28070	3.45833	P > .95	Accept
4. Night Driving Fatigue	P .05	3.72368	3.91203	P > .95	Accept
5. Proper Headlight Use	P< .05	4.30000	4.12963	P > .95	Accept
6. Pedestrians	P< .05	4.22368	3.99537	P > .95	Accept
7. Driving at Night	P< .05	3.52631	3.60185	P > .95	Accept
8. Total Concepts	P< .05	3.91040	3.94378	P > 95	Accept
		Mean Scores II	**Mean Scores IV**		
Permanent Attitude Shifts					
1. Night Driving Skill	P< .05	4.62500	5.37500	P < .95	Reject
2. Headlight Limitations	P< .05	4.52679	4.40385	P > .95	Accept
3. Night Driving Visibility	P< .05	3.41964	3.28846	P > .95	Accept
4. Night Driving Fatigue	P< .05	3.54464	4.27885	P < .95	Reject
5. Proper Headlight Use	P< .05	4.09821	3.99039	P > .95	Accept
6. Pedestrians	P< .05	4.27678	4.18269	P > .95	Accept
7. Driving at Night	P< .05	3.60714	3.58654	P > .95	Accept
8. Total Concepts	P< .05	4.01403	4.15797	P > .95	Accept

TABLE III

RESULTS OF THE F TESTS, t-TESTS, MEAN SCORES, DEGREES OF FREEDOM, AND N FOR CHANGE IN DRIVING PERFORMANCE

F Test	t-Test	N		Mean Scores		Degrees of Freedom
		I	II	I	II	
P <. 05	P >. 95	49	28	5. 2245	5. 4643	75
Analyses of Groups Divided According to Skill						
		Skilled-Normal		Skilled	Normal	
P <. 05	P >. 95	90	54	5. 5625	5. 3148	148
Matched Group Analyses, I and III, II and IV						
		I	III	I	III	
P <. 05	P >. 95	47	48	5. 2245	5. 8542	95
		II	IV	II	IV	
P <. 05	P >. 95	28	25	5. 46429	5. 2400	51

nent tested as though it were a separate problem. However, the acceptance or rejection of the original hypothesis was based upon the degree of conformance of all three of the components to the conditions set forth in the study.

First, the writer determined the correlation between the immediate learning estimates of the panels as to the proportion of total students answering questions correctly after, but not before, the film viewing, and the difference between the proportion of correct answers to an item by Control Group I and Experimental Group III. This process was then repeated for Control Group II and Experimental Group IV to show the correlation of the retention estimates.

Since all groups used the same test, the difference between the scores made by the students on each test item was used in the statistical analysis. These differences were correlated with the predictions for immediate change made by the two panels of experts representing both subject matter (Traffic Safety Panel) and film production experts (Audio-Visual Panel). The Pearson r, coefficient of linear correlation, was used because of its ability to determine the mutual relationship between two variables such as predicting of achievement from a prognostic test (4, p. 233).

The Pearson r was applied to the following relationships:

1. Correlation between the predictions of the two panels of experts.
2. Correlation between the combined group predictions of the expert panels and the population results.
3. Correlation between the predictions of the Audio-Visual Panel and the population reresults.
4. Correlation between the predictions of the Traffic Safety Panel and the population results.

Each of these correlations was then tested by the t-test of significance as a means of later accepting or rejecting the hypothesis.

In general, the positive relationships found in the above correlations appear to be significant and to reveal substantial agreement between the predictions of the various panels and the results obtained by the testing of sample subjects for learning achievement and retention. It is evident that the agreement between the two variables tested in each case is much greater than a chance relationship. On the basis of this evidence, the writer would accept the seventh hypothesis as far as the immediate and retention effects of the first component (learning) are concerned.

The second component of the "behavioral analysis" required the writer first to correlate the predictions of the two panels of experts in regard to the immediate change in attitudes anticipated as a result of viewing the film. Next, the writer determined the correlation between the estimates of the panels for all seven concepts for a shift in the desired direction and the actual desired direction shift observed in the sample population (based upon the difference between Control Group I and Experimental Group III). This same process was repeated for Control Group II and Experimental Group IV to show correlations in the retention of attitude shifts.

All groups used the same attitude test, and the differences between the scores made by the students on each concept of the test was utilized in the statistical analysis. The Pearson r coefficient was applied to the following relationships:

1. Correlation between the predictions made by the two panels of experts as to immediate attitude shift.
2. Correlation between the immediate attitude predictions of the Traffic Safety Panel and the test results.
3. Correlation between the immediate attitude predictions of the Audio-Visual Panel and the test results.
4. Correlation between the combined group (Traffic Safety and Audio-Visual) immediate attitude predictions and the test results.

Upon completion of the above correlations, the t-test of significance of the value obtained for r was applied to the correlation between the two panels of experts. The significance for the other r values was determined by use of a table designed to show this relationship.

Although the relationships found in the immediate and retention attitude prediction correlations were positive, they were too small and did not have substantial enough agreement to be considered significant in this study. This evidence would tend to reject the hypothesis with reference to the predictability of immediate and permanent attitude shifts.

In this experiment, the term "driving skill" refers to potential driving behavior or performance when performing a specific driving function. This component of the behavioral analysis intends to measure the correlation between the predictive estimates of the expert panels and the actual results obtained when the students were administered a driving test. It specifically attempts to determine the correlation between the estimates of the increase in proficiency on nine items, and the difference between Control Group I and Experimental Group III. Upon completion of the analysis of these groups, the process is then repeated for Control Group II and Experimental Group IV.

All of the groups used the same driving tests. However, Groups I and III were administered the test immediately after viewing the film in order

TABLE IV

COEFFICIENTS OF LINEAR CORRELATION AND t-VALUES OF SIGNIFICANCE FOR ALL COMPONENTS OF THE BEHAVIORAL ANALYSIS

Component	r_{12}	t	r_{14}	t	r_{24}	t	r_{34}	t
Immediate Learning	.5189	$t_{.01}$	.2175	$t_{.10}$	.4303	$t_{.01}$	.3571	$t_{.05}$
Permanent Learning	.2913	$t_{.10}$	.2575	$t_{.10}$	.3479	$t_{.05}$	.4753	$t_{.01}$
Immediate Attitude Change	.9280	$t_{.01}$	.3003	$t_{.10+}$	.3611	$t_{.10+}$	.3367	$t_{.10+}$
Permanent Attitude Change	.7757	$t_{.05}$	.0602	$t_{.10+}$	.1192	$t_{.10+}$	.0914	$t_{.10+}$
Immediate Skill Change	-.2116	$t_{.43}$	.3112	$t_{.10}$	-.053	$t_{.40}$	.392	$t_{.50}$
Permanent Skill Change	-.2065	$t_{.50}$	.1128	$t_{.50}$	-.5649	$t_{.10}$	-3.3787	$t_{.40}$

r_{12} = Traffic Safety Predictions vs. Audio-Visual Predictions
r_{14} = Traffic Safety Predictions vs. Population Scores
r_{24} = Audio-Visual Predictions vs. Population Scores
r_{34} = Combined Group (T.S. and A.V.) Predictions vs. Population Scores

to observe the immediate film effects, and Groups II and IV received the test after six weeks' delay in order to determine the permanency of the film effect upon driving skill. The differences between the scores made by the students on each of the nine items tested were used in the statistical analysis. The differences in scores between the experimental and control groups were correlated with the panel predictions by using the Pearson r, coefficient of linear correlation, which was applied as follows:

1. Correlation between the predictions of the two panels of experts.
2. Correlation between the combined group predictions of the expert panels and the population results.
3. Correlation between the predictions of the Audio-Visual Panel and the student results.
4. Correlation between the predictions of the Traffic Safety Panel and the student results.

Upon the completion of each correlation, each r-value was tested by the t-test of significance as a basis for accepting or rejecting the hypothesis.

The relationships observed in the above correlations showed little agreement between the variables measured and cannot be considered significant in this experiment. The evidence also tends to reject the major hypothesis with reference to the predictability of change in immediate driving performance.

Table IV lists the r coefficients of linear correlation and the t-values of significance for all of the component areas of the behavioral analysis. These correlations reveal high predictability for increase in learning, and low predictability correlation for both attitude and behavior change. The seventh hypothesis, which states that it is possible tp predict, by the use of a behavioral analysis, the measurable change in knowledge, driving attitudes, and driving skills which will occur as a result of viewing a traffic safety film, is so stated that its acceptance or rejection must be based upon the conformance of all three components to the conditions set forth in the study. Therefore, on the basis of the total evidence, the writer rejected the hypothesis.

Summary

The purpose of the study was to obtain data to test the hypotheses which were set up regarding the effectiveness of a traffic safety motion picture film upon its viewers. This involved determining the immediate and permanent increase in learning, the immediate and permanent shift in attitudes, the immediate and permanent change in driving skill, and the development of a predictive instrument, called "Behavioral Analysis" by which these changes might be anticipated.

The film "Driving At Night" was selected as the experimental film after thirty traffic safety films had been previewed. Factors such as the frequency of use of the film in driver training courses, the educational level for which the film was best suited, the subject matter content, the quality of photography and sound effects, and the adaptability of the film to the objectives of this particular research were factors in the film selection. A behavioral analysis was developed by projecting the film and making a record of all of the important commentary and visual events as individual statements of fact. After the facts in the film were categorized, the reaction of the viewer was analyzed by listing that information which we expected the learner to associate with the fact. At this point, estimates were made of the impact of the film. Next, the writer analyzed potential behavioral change into its components. The components were obtained from personal interviews and street corner observations. Each one of the behavior change components was analyzed to determine how behavior might change. Upon the completion of the "behavioral analysis", it was administered to ten judges consisting of five Traffic Safety experts and five Audio-Visual experts, who made predictions on possible changes in learning, attitude, and driving performance which might result from viewing the film.

The data to be obtained from the sample population included measurement of achievement, attitude shift, and change in driving skill. The writer constructed and pretested instruments to measure achievement and attitude shift. The reliability of the achievement test was established at $r = +.93$ with $N = 37$. It was tested and found to be significant at $t_{.001}$ with 35 degrees of freedom. Reliability of the attitude test was found to be +98 for the seven concepts, and $r = +.93$ with $N = 49$ for the items. These correlations were revealed to be significant at $t_{.001}$ with 47 degrees of freedom. The skill test was not pretested but was constructed in such a manner as to allow direct observation of the driving behavior of each member of the sample population.

In the experiment, emphasis was placed on the ability to estimate and obtain effects which were due only to the strength of the film and not by how it was administered. Therefore, the verbal instructions were short, and the written instructions were designed to enable the student to understand what was required on each test. May and Lumsdaine state that the best results are obtained and learning is increased by: 1) use of the film at the correct ability level of the students, 2) using the film in context with the subject taught and the unit of instruction, 3) giving prior instruction on the content of the film and the purpose

for seeing it, 4) stopping the film for practice or test questions, and 5) clarifying obscure points in the film after it is viewed (1, p. 310). They also indicate that "a strong interaction effect will be found between these divices and the general competence of the teacher" (1, p. 317). Therefore, the researcher, recognizing the important roles of the teacher and supplementary instruction limits his general conclusions to those findings which may represent film effects obtainable when a film is used out of phase with any unit of instruction.

The groups for the study were drawn from a population of sophomore, junior, and senior students at Michigan State University. The selection of these people was by both the purposive and random assignment methods. The experimental design placed (by random selection) all members of the sample population into one of two control or or two experimental groups. Control Groups I and II viewed a non-related film and Experimental Groups III and IV viewed the experimental film. Control Group I and Experimental Group III were tested immediately after viewing the films for immediate effects. Control Group II and Experimental Group IV were tested six weeks after viewing the film in order to determine the permanent effects.

The F-test was employed to determine if the variances of the two populations might be presumed equal in the analysis of the data obtained concerning the six null hypotheses. The t-test of significance was used to determine whether the results obtained were significant. The Pearson r coefficient of linear correlation was employed in the analysis of the data obtained for purposes of evaluating the behavioral analysis technique. The r-value was then tested by the t-test of significance in order to determine whether to accept or reject the seventh hypothesis.

In view of the fact that there was statistically significant difference found in achievement between Groups I and III, and Groups II and IV, it was logical to reject both the first and second null hypotheses. There was no statistically significant difference in attitude shift or change in driving skill between Groups I and III, and Groups II and IV. Therefore, the writer accepted the third, fourth, fifth, and sixth null hypotheses. The three basic components of the "Behavioral Analysis" revealed statistically significant difference in regards to the prediction of learning, but no statistically significant difference with reference to the predictions of change in attitude shift and driving skill. Therefore, the seventh hypothesis was rejected because of the lack of positive results for all three components.

The results of the statistical analyses reveal that experienced drivers do learn factual information from viewing traffic safety films, and this learning is dominant both immediately after viewing the film and for a long period afterward. However, the learning of factual information does not necessarily coincide with a corresponding shift in attitudes or improvement of driving skill. The analysis also reveals that it is possible to predict the proportion of increase in learning which may be obtained from viewing a film, but that it is very difficult to predict shifts in attitude and change in driver performance. It is possible that the failure to predict these two components successfully is due to the small amount, if any, of change that actually occurs as a result of viewing the film.

Careful effort was made in the experiment by the researcher to control all variables which could be anticipated. Special tests were constructed which were used as both pretests and final tests, and the tests were administered under similar environmental conditions in order to prevent contamination by unforseen variables. All of the tests were submitted to subject matter experts and driver training teachers for their evaluation and suggestions. Those factors which could not be controlled were randomized, so that their effects could enter into the estimate of error term provided in the statistical analysis.

Therefore, on the basis of the type of experiment used, the kind of analyses made, and on the evidence found, it seems logical to infer that motion pictures concerning Traffic Safety make significant contributions to learning, but little contribution to attitude shift or improved driving performance, insofar as the conditions of use were those followed in this experiment. Upon the completion of the film analysis, the writer compared the deviant cases, resulting from bad predictions by the panel of experts, with the type of film treatment given these particular points of instruction. This comparison revealed no differences in film treatment which might explain why these particular cases were deviant.

Conclusions

It would seem to the writer that the findings of this experiment further emphasize the need for many carefully designed and statistically analyzed studies before a generalization can be made concerning all Traffic Safety films. The results of the evaluation of this film are indicative of the benefits and effects which may be obtained by a teacher who makes only casual use of films as a supplement to a learning situation. In reviewing the steps taken in conducting the research, the writer finds a generally systematic approach to the study of the problems involved. However, fault may be found in the administration of the

"Behavioral Analysis" to the two panels of experts. Possibly significant results might be obtained if a replicate of the study were to include a more elaborate orientation of the judges to the predictive instrument. This could result in a more systematic evaluation by the judges of all subjective ratings of the behavioral change components.

The findings of this experiment lead the writer to believe that the predictive instrument may be used equally as well by the subject matter (Traffic Safety Panel) or film production (Audio-Visual Panel) experts. The subject matter experts did slightly better in predicting driver change, slightly worse in the prediction of learning, and about equally well in the prediction of attitudes when these factors were correlated against student performance. The differences were generally not significant and a very high degree of correlation was obtained when the two expert panels were compared to each other.

The writer would recommend that further study be made concerning the following:

1. Apply the predictive instrument to several films to determine its ability to predict accurately the total amount of learning, attitude shift, and skill improvement which may result from viewing a film.
2. Investigate all deviant estimates where both panels were about equally deviant.
3. Investigate all estimates where one panel estimated closely and where the other panel was off.
4. Determine if any relationships exist between deviant estimates and the type of information presented concerning that portion of the film.
5. Investigate the relationship between the use and success of the "Behavioral Analysis" and the purposes of the film content.
6. Investigate the possibility of producing future Traffic Safety films based upon the behavior change components listed in the "Behavioral Analysis."
7. Investigate further the cause and nature of the "sleeper effect" found in the tests for attitude shift, and determine whether or not this same type of effect may also exist in tests of learning and behavior.

REFERENCES

1. May, Mark A., and Lumsdaine, Arthur A., Learning From Films (New Haven: Yale University Press, 1958).
2. Osgood, Charles E., Suci, George J., and Tannenbaum, Percy H., The Measurement of Meaning (Urbana, Illinois, University of Illinois Press, 1957).
3. U. S. Army and U. S. Navy, Instructional Film Research, 1918-1950. NAVEXOS P -977. (New York: U. S. Naval Training Device Center, Port Washington, Long Island, New York, 1956).
4. Walker, Helen, and Lev, Joseph, Statistical Inference (New York: Henry Holt and Co. 1953).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 1, September 1962)

A STUDY OF GRADUATES OF THE UNIVERSITY OF GEORGIA WHO ARE CERTIFIED TO TEACH WITH RESPECT TO ENTRANCE INTO THE TEACHING PROFESSION

MARY OELLERICH MURPHY
Madison College, Harrisonburg, Virginia

THERE HAS never been an oversupply of highly competent teachers in our public schools. There is at present a shortage of fully qualified teachers. Only a part of those needed are being prepared by our colleges and universities.[1] Some of those who are trained as teachers never teach.

The two basic ingredients of an adequate system of education are an appropriate curriculum and qualified teachers. Neglect of either results in damage to the total enterprise.

How to attract and hold an adequate supply of well-qualified teachers is a major concern of our society today. An undisputed cause and effect relationship is that from the quality of the teacher flows directly the quality of the education the child receives.

One index of a true shortage is the number who leave teaching to enter other occupations. An equally important index is the number who train for teaching but who fail to enter the profession.

It is a primary responsibility of a democratic society to recruit and to prepare teachers of the highest quality. The degree to which our quality of living is improved is directly related to the extent to which this responsibility is effectively assumed and executed.

If highly competent teachers are to be effectively recruited, it is important to determine and understand factors associated with those who are trained to teach and who actually enter the profession. The task of developing potentially outstanding teachers and of guiding them into teaching as a vocation is essentially a problem with which colleges and universities should be greatly concerned. To be effective in guiding and counseling future teachers, educators need to understand the differences between graduates trained as teachers who teach and graduates trained as teachers who fail to teach.

Problem

The problem of the present study is to compare a group of graduates who trained as teachers and who subsequently entered the teaching profession with a group of graduates who trained as teachers but who did not enter the teaching profession. This comparison will be made on the basis of the following groupings of factors: 1) The personal background of the graduate, 2) the secondary school background of the graduate, 3) the scores made by the graduate on placement examinations administered by the University of Georgia, and 4) the college performance of the graduate.

Hypothesis

The hypothesis of this study is that there are significant differences between those graduates trained as teachers who teach and those graduates trained as teachers who fail to teach.

Population

The June graduates of the University of Georgia for 1957 and 1958 were selected as the population on the basis of test results available, recency of attendance, and adequacy of information available. A total of 496 graduates qualified as trained teachers[2] at the baccalaureate level. Of 496 graduates, 138 did not enter the teaching profession and 324 did enter the teaching profession. These two groups comprised the base population for this study. The desired information could not be obtained for 34 graduates.

---

1. C. R. Spain, "Personnel: Salaries and Benefits Bolstered but Teacher Shortage Still Acute," School Executive, 78:66-67, January, 1959.
2. "Trained teachers" means those graduates qualified to receive the four-year professional teacher's license from the Georgia State Department of Education.

Statistical Analyses

Two types of statistical analyses were used in this study to compare the two groups. The chi-square technique was used to compare the total distributions of the two groups. The "t" test was used to compare the means of the two groups.

The chi-square technique was employed to test the hypothesis that the differences in the distributions are due to chance alone. When the operation of chance factors alone could be considered to account for the observed difference in the two distributions fewer than five times in one hundred (the five percent level of confidence), the tendency observed was considered to be the result of other than chance operation; that is, the two groups being compared had a "significant" difference in the two distributions.

For the continuous variables, such as test scores, the "t" test was employed to test the hypothesis that the mean scores of the two groups were equal; that is, the difference between the means was due to chance alone. When the operation of chance alone could be considered to happen less than five times in one hundred (the five percent level of confidence), the tendency observed was considered to be the result of other than chance operation; that is, a significant difference existed between the mean scores of the two groups under consideration.

Results

The following is a listing of the 46 individual factors under consideration, together with the significance (at the five percent level of confidence or better) or lack of significance for each factor when the mean scores of the two groups were compared and when the overall distributions of the two groups were compared.

Factor	Significance of Means	Significance of Distributions
A. Personal Background Factors		
1. Continuous factors		
a. Age	Yes*	No
2. Non-continuous variables		
a. Place of birth	--	No
b. Sex	--	Yes***
c. Religious preference	--	No
d. Size of city of residence	--	No
e. State of residence	--	No
f. Father living	--	No
g. Mother living	--	No
h. Parents separated (other than by death)	--	No
i. Parents remarried or guardian	--	No
j. Occupation of father	--	No
k. Occupation of mother	--	No
B. Secondary School Background Factors		
1. Continuous factors		
a. Number of secondary schools attended	No	No
b. Number of English credits earned	No	No
c. Number of foreign language credits earned	No	No
d. Number of mathematics credits earned	No	No
e. Number of science credits earned	No	No
f. Number of commercial credits earned	No	No
g. Number of social science credits earned	No	No
h. Number of academic credits earned	No	No
i. Number of activity and vocational credits earned	No	No
j. Total number of credits earned	No	No
2. Non-continuous variables		
a. Type of control (public, private, parochial)	--	No

* Significant at the five percent level of confidence.
** Significant at the one percent level of confidence.
*** Significant at the point one percent level of confidence.

Continued. . . .

C. Freshman Placement Examination Factors		
1. Continuous factors		
a. ACE "Q"-score	No	No
b. ACE "L"-score	No	No
c. ACE "T"-score	No	No
d. Algebra Placement Examination score	No	No
e. Cooperative Reading "vocabulary" score	No	No
f. Cooperative Reading "speed of comprehension"	No	No
g. Cooperative Reading "level of comprehension"	No	No
h. Cooperative Reading "total" score	No	No
D. College Performance Factors		
1. Continuous factors		
a. Number of years between high school graduation and baccalaureate graduation	Yes**	Yes*
b. Actual average for work at University	No	No
c. Graduating average for work at University	No	No
d. Total number of quarter hours in Education	No	No
e. Total number of quarter hours in teaching field	No	No
f. Number of probations	No	No
g. Number of suspensions	No	No
h. Number of colleges attended before entering the University	No	No
i. Number of "D" grades	No	No
j. Number of "F" and "WF" grades	No	No
k. Number of quarter hours in undergraduate program	No	No
2. Non-continuous factors		
a. Teaching field	--	Yes***
b. Type system from which entered	--	No
c. Degree program for which registered	--	Yes***
d. Degree program from which graduated	--	Yes***

The following paragraphs are interpretations of the six individual factors found to differentiate significantly between graduates trained as teachers who entered the teaching profession and graduates trained as teachers who did not enter the teaching profession. The reader is referred to Tables I - VI for a more detailed statistical analysis of all factors in this study.

Age. The "t"-test indicates that the mean age of the entries into teaching is significantly less than the mean age of the non-entries. Although the chi-square test indicates a tendency for the entries to be associated with a younger age, this tendency is not reliable at the five percent level of confidence or better. A larger percentage of the entries than of the non-entries was 21 years of age or younger (43.8 percent versus 39.9 percent). A small percentage of the entries than of the non-entries was 26 years of age or older (12.3 percent versus 20.3 percent).

Sex. More women certified to teach entered the teaching profession than would be expected from chance alone, the percentage being 72.8 for the women entries versus 27.2 for the men. A larger percentage of the entries than of the non-entries was of women (72.8 percent versus 54.7 percent). Conversely, a smaller percentage of the entries than of the non-entries was men (27.2 percent versus 45.3 percent).

Number of years between high school graduation and baccalaureate graduation. The mean number of years for the entries was significantly less than that for the non-entries (5.72 years for the entries versus 7.14 years for the non-entries). The same pattern was observed for the distributions of frequencies.

Teaching field. Larger percentages of entries than non-entries were in the teaching fields of business education, elementary education, and physical education (4.3 percent versus 2.9 percent, 37.3 percent versus 23.4 percent, and 14.2 percent versus 7.3 percent, respectively). Smaller percentages of entries than non-entries were in the teaching fields of agricultural education, English, and social studies (5.6 percent versus 16.8 percent, 6.2 percent versus 7.3 per-

TABLE I

SUMMARY OF COMPARISONS OF ENTRIES AND NON-ENTRIES BASED ON CONTINUOUS PERSONAL BACKGROUND FACTORS

Factor	N	Entries		N	Non-Entries		$M_E - M_N$	"t"	"p_t"	"p_{x^2}"
		Mean	S. D.		Mean	S. D.				
Age in months	324	284.13	65.25	138	298.50	70.55	14.37	2.05	$<.05$*	$<.10$

* Significant at the five percent level of confidence.
\> Greater than.
< Less than.

TABLE II

SUMMARY OF COMPARISONS OF ENTRIES AND NON-ENTRIES BASED ON NON-CONTINUOUS PERSONAL BACKGROUND FACTORS

Factor	"p_{x^2}"
Place of birth	>.20
Sex	<.001***
Religious preference	<.80
Size of city of residence	>.70
State of residence	<.30
Father living	1.00
Mother living	1.00
Parents separated (other than by death)	<.30
Parents remarried or guardian	1.00
Occupation of father	>.50
Occupation of mother	>.10

***Significant at the .1 percent level of confidence.

TABLE III

SUMMARY OF COMPARISONS OF ENTRIES AND NON-ENTRIES BASED ON CONTINUOUS SECONDARY SCHOOL BACKGROUND FACTORS

Factors	N	Entries Mean	Entries S. D.	N	Non-Entries Mean	Non-Entries S. D.	$M_E - M_N$	"t"	"p_t"	"$p_x{}^2$"
Number of secondary schools attended	323	1.14	.37	138	1.13	.35	.01	.28	$<.80$	1.00
Number of English credits earned	305	4.00	.34	132	3.96	.44	.04	.97	$>.30$	$>.50$
Number of foreign language credits earned	305	1.43	1.37	132	1.44	1.40	.01	.07	$>.90$	$<.98$
Number of mathematics credits earned	305	2.77	.86	132	2.89	.83	.12	1.38	$<.20$	$<.70$
Number of science credits earned	305	2.01	.85	132	2.08	.90	.07	.77	$>.40$	$<.70$
Number of commercial credits earned	305	1.20	1.13	132	.95	1.02	.25	.72	$<.50$	$>.10$
Number of social science credits earned	305	2.72	.90	132	2.69	.82	.03	.40	$<.70$	$>.50$
Number of academic credits earned	305	13.50	5.24	132	12.97	2.15	.53	1.50	$>.10$	$>.95$
Number of activity and vocational credits	305	3.71	1.94	132	3.77	2.13	.06	.30	$<.80$	$<.30$
Total number of credits earned	302	16.84	.94	133	16.80	1.37	.04	.36	$>.70$	$<.10$

Note: The reader is reminded that it is permissible to treat non-continuous variables as if they were continuous.

TABLE IV

SUMMARY OF COMPARISONS OF ENTRIES AND NON-ENTRIES BASED ON CONTINUOUS FRESHMAN PLACEMENT EXAMINATION FACTORS

Factor	N	Entries		N	Non-Entries		M_E-M_N	"t"	"p_t"	"$p_x{}^2$"
		Mean	S. D.		Mean	S. D.				
ACE "Q"-score	112	41.98	9.64	41	39.66	9.33	2.32	.80	$<.40$	$<.70$
ACE "L"-score	112	56.56	15.23	41	60.20	14.64	3.64	.50	$>.60$	$<.50$
ACE "T"-score	110	97.86	29.02	41	99.98	21.68	2.12	.48	$>.60$	$>.50$
Algebra Placement Examination score	110	25.26	9.99	41	27.10	13.69	1.84	.33	$>.70$	$>.50$
Cooperative Reading "vocabulary" score	105	33.74	13.96	40	35.80	15.00	2.06	.71	$<.50$	$>.05$
Cooperative Reading "speed of comprehension"	105	33.91	16.24	40	36.13	15.37	2.22	.76	$>.40$	$<.10$
Cooperative Reading "level of comprehension"	108	46.47	14.60	41	47.32	13.45	.85	.33	$>.70$	$>.10$
Cooperative Reading "total" score	105	151.30	27.75	39	156.87	26.25	5.57	1.11	$<.30$	$>.05$

TABLE V

SUMMARY OF COMPARISONS OF ENTRIES AND NON-ENTRIES BASED ON CONTINUOUS COLLEGE PERFORMANCE FACTORS

Factor	N	Entries		N	Non-Entries		$M_E - M_N$	"t"	"p_t"	"p_{x^2}"
		Mean	S. D.		Mean	S. D.				
Number of years between high school graduation and baccalaureate graduation	324	5. 72	4. 69	138	7. 14	6. 00	1. 42	2. 49	$>$. 01**	$<$. 05*
Actual average for work at University	321	82. 68	3. 61	136	82. 86	6. 33	. 18	. 32	$>$. 70	$<$. 80
Graduating average for work at University	321	82. 73	4. 47	136	83. 34	5. 66	. 61	1. 12	$<$. 30	$<$. 70
Total number of quarter hours in Education	321	42. 09	8. 12	136	41. 70	8. 60	. 39	. 30	$<$. 80	$>$. 80
Total number of quarter hours in teaching field	322	56. 75	17. 66	136	58. 06	20. 76	1. 31	. 59	$<$. 60	$>$. 30
Number of probations	322	. 09	. 41	136	. 13	. 57	. 04	. 85	$<$. 40	$<$. 70
Number of suspensions	322	. 02	. 17	136	. 01	. 12	. 01	. 52	$>$. 60	1. 00
Number of colleges attended before entering the University	324	. 75	. 69	137	. 80	. 70	. 05	. 74	$<$. 50	$>$. 70
Number of "D" grades	322	2. 17	2. 62	136	2. 34	2. 81	. 17	. 59	$<$. 60	$>$. 90
Number of "F" and "WF" grades	322	1. 20	2. 28	136	1. 10	2. 07	. 10	. 47	$>$. 60	$>$. 50
Number of quarter hours in undergraduate program	322	202. 03	12. 21	136	202. 15	12. 41	. 12	. 10	$>$. 90	$<$. 20

* Significant at the five percent level of confidence.
** Significant at the one percent level of confidence.

cent, and 6.2 percent versus 16.8 percent, respectively). A larger percentage of entries than non-entries was in all other teaching fields not listed above when these fields were treated as a combined group (14.5 percent versus 13.9 percent). An equal percentage of entries and non-entries was in the teaching field of home economics (11.7 percent versus 11.7 percent). Fields not mentioned contained such small quantities as to make a comparison invalid.

Degree program for which registered. A larger percentage of entries than non-entries registered for the degree program of B. S. Degree in Education (68.3 percent versus 47.8 percent). This was the only degree program in which the percentage of entries exceeded the percentage of non-entries. Smaller percentages of entries than non-entries registered for the degree programs of 1) A. B. (9.2 percent for entries versus 10.9 percent for non-entries), 2) B.S. (3.7 percent for entries versus 5.8 percent for non-entries), 3) B. S. Degree in Home Economics (8.9 percent for entries versus 11.6 percent for non-entries), 4) B. S. Degree in Agriculture Engineering (5.5 percent for entries versus 18.1 percent for non-entries), and 5) all other degree programs combined into one group (4.3 percent for entries versus 5.8 percent for non-entries).

Degree program from which graduated. A larger percentage of entries than non-entries graduated from the degree program of B. S. Degree in Education (83.1 percent versus 71.1 percent). This was the only degree from which graduated in which the percentage of entries exceeded the percentage of non-entries. Smaller percentages of entries than of non-entries graduated from the degree programs of 1) B. S. Degree in Home Economics, and 2) B. S. Degree in Agriculture and B. S. Degree in Agriculture Engineering (11.3 percent versus 11.9 percent and 5.6 percent versus 17.0 percent, respectively).

Conclusions

The data presented have established, as originally hypothesized, that certain significant differences do exist between graduates trained as teachers who subsequently entered the teaching profession and graduates trained as teachers who did not enter the teaching profession. Certain of the basis of intergroup comparison failed to differentiate significantly between the two groups.

Potential entries into the teaching profession are most likely to be found among individuals who have registered for the degree program of B. S. Degree in Education and who have graduated from that program. It is probable that persons entering teaching think that the combination of content courses and professional courses offered for this degree best qualify them for the duties of a teacher.

The larger percentages of trained teachers who enter the profession are most likely to be found in the teaching fields of business education, elementary education, and physical education. Entries into the teaching profession are less likely to be found in the teaching fields of agriculture education, English, and social studies. The entries and non-entries into teaching from the field of home economics are likely to be of equal probability. The number of trained teachers in the teaching fields not mentioned above were too small in number to make a valid comparison in this study. The entries and non-entries from all other teaching fields, combined into one group, are likely to be of equal probability. It is probable that persons trained in the teaching fields of business education, elementary education, and physical education have more limited opportunity for employment outside of the field of teaching.

Potential entries into the teaching profession are more likely to be found among women than among men. Approximately 73 percent of the entries can be expected to be women and 27 percent can be expected to be men. Some of this difference may well be due to the small salaries for men in the teaching profession as compared to other professions and other business opportunities. Also, employment opportunities for women may be somewhat less than those for men. The fact that many women trained to teach do not enter the profession because of marriage and child-rearing appears to have little effect upon the above-mentioned percentages. This may well be due to a more favorable attitude toward the working wife and working mother.

Potential entries into teaching can be expected to be more than one year and six months younger than non-entries. More entries than non-entries can be expected to be twenty-one years old and younger. Fewer entries than non-entries can be expected to be twenty-six and older. The younger age of entries as compared to that of non-entries into teaching may be accounted for by the fact that a smaller percentage of men are entries than are non-entries. Many men have their college education interrupted by military service. Since a smaller percentage of men are entries than are non-entries, the total effect may be to increase the age for the group of non-entries as a whole.

Potential entries into teaching can be expected to have significantly fewer years between high school graduation and baccalaureate graduation as compared to non-entries. About one-half as many entries as non-entries can be expected to have nine and more years between high school graduation and baccalaureate graduation. Some

TABLE VI

SUMMARY OF COMPARISONS OF ENTRIES AND NON-ENTRIES BASED ON NON-CONTINUOUS COLLEGE PERFORMANCE FACTORS

Factor	"p_{x^2}"
Teaching field	<. 001***
Type system from which entered	>. 05
Degree program for which registered	<. 001***
Degree program from which graduated	<. 001***

***Significant at the . 1 percent level of confidence.

of this can no doubt be accounted for by the fact that a smaller percentage of men are entries than are non-entries. Many men have their college education interrupted by military service. Also, many of these men are provided the opportunity of a college education after military service through funds provided by the government. Since a smaller percentage of men are entries than are non-entries, the total effect may be to lengthen the number of years between high school graduation and baccalaureate graduation for the group of non-entries as a whole.

Secondary school background factors and freshman placement examination factors do not differentiate between graduates trained as teachers who enter the teaching profession and graduates trained as teachers who do not enter the teaching profession. It appears on the bases of these results that the minimum entrance requirements as set up by the University are sufficient indications of the academic ability required of teachers. The present data have been interpreted to indicate that motivational factors are highly important determiners of the probability of entrance into teaching. Academic ability as measured by standardized tests such as the ACE and credits earned in secondary school are not the yardsticks by which potential entries into teaching are to be evaluated.

CONFIGURATIONAL INVARIANCE IN THE CALIFORNIA PSYCHOLOGICAL INVENTORY

JOHN PIERCE-JONES, JAMES V. MITCHELL, JR.
University of Texas
F. J. KING
Florida State University

IN AN EARLIER paper, two of the present authors (Mitchell and Pierce-Jones, 11), reported a factor analytic investigation of the 18 scales of the California Psychological Inventory (CPI) for a sample of 258 university students. After orthogonal rotations had been performed by Kaiser's (10) varimax method, four factors were provisionally named as follows: 1. Adjustment by Social Conformity; 2. Social Poise or Extraversion; 3. Superego Strength; and 4. Capacity for Independent Thought and Action. The earlier study was prompted by essentially practical questions surrounding the interpretability of CPI profiles in view of the substantial correlations among the CPI scales. Although he brought no empirical evidence to bear on such practical matters, Becker (4) recently observed that the 18 CPI variables probably could be reduced to four or five essentially uncorrelated measures, a judgment which our earlier results tended to support.

Problem

The replication of factor analytic studies on descriptively different populations is a sine qua non for the intelligent evaluation of the characteristics and consistency of the factorial pattern revealed. It should not be assumed that factor loadings are always fixed and immutable realities that are invariably of the same order of magnitude regardless of the nature of the sample employed. In fact, it is well established that factor loadings for the same test battery will differ rather widely for groups of divergent education, background, and experience. The fundamental question is not whether the loadings of a given test battery will differ from one sample to another, but whether they will differ to such an extent as to require modification in factor interpretation or the recognition of fundamental differences between the factors represented in the separate samples. Thurstone recognized this clearly when he discussed the liklihood of "configurational invariance" in the personality domain (14, p. 365). According to Thurstone, "configurational invariance" was established if it could be demonstrated that factor identities were fundamentally the same for different samples of subjects who had taken the same test battery, and such invariance was considered a rather crucial issue in personality research.

The present study attempts to apply this important criterion of "configurational invariance" to the factorial structure of the California Psychological Inventory. The implementation of this goal was accomplished by structuring the present investigation so as to fulfill two specific purposes: a) to perform, by exactly the same methods as used in our earlier study (11) a factor analytic investigation of the CPI scores obtained for a sample of a population which was descriptively different from that employed earlier; and b) to assess the configurational invariance of personality factors located in this and our earlier study by the method of "transformation analysis" developed, applied, and described by Ahmavaara (1, 2). Succinctly put, the present investigation was directed at evaluating the configurational invariance of personality factors across samples from different populations.

Procedure

Subjects. In the earlier research, already cited, with which the present study was also necessarily concerned, 258 university undergraduate students (213 females, 45 males) were employed as subjects. In the present study, CPI scores were obtained for 156 male public school superintendents—80 percent of the chief city

school administrators in a single state.[1]* This sample has been described comprehensively by Carleton (5), who concluded that these men were probably representative of all city school superintendents in the state. In addition to their occupational homogeneity, behind which existed marked similarity of academic and professional experience, these men were much older (mean CA = 43.5 years) than the mainly female student sample used in our original analysis of the CPI.[2]

Factorization and Rotation. A matrix of the 153 product-moment correlations among the 18 CPI scales was constructed for the school administrators. It yielded four factors when subjected to centroid analysis (Guilford, 9). Application of Humphry's rule (Fruchter, 7) and Burt's empirical formula (Thomson, 13) led to termination of factorization after extraction of the fourth factor. Analytic orthogonal rotations were then made by means of an IBM Model 650 electronic computer, using Kaiser's varimax method (10). These procedures exactly duplicated those used in the original study by Mitchell and Pierce-Jones (11).

Transformation Analysis. To afford better than an intuitive comparison of the results obtained in this and our earlier CPI research, Ahmavaara's method of "transformation analysis" was applied (1, 2). The mathematical method of Tucker (15) might have been used, perhaps, but we were unaware of its having been so extensively or systematically employed to compare independent factor analytic studies as has transformation analysis (2).[3]

Explicit instructions for transformation analysis are readily available (2, pp. 45-58). What is essentially involved is the computation of a transformation matrix, L, using the two final factor matrices, F_I and F_{II}, from the separate factor studies according to the formula:

$$L = (F_I'F_I)^{-1}F_I'F_{II}$$

This transformation matrix contains invariance values or the loadings of the factors of Study II (present study) on the factors of Study I (11).[4] "Every row of matrix L refers to a factor of Study I, every column to a factor of Study II, the factors being taken in the same order as... in the matrices F_I and F_{II} respectively" (2, pp. 45-46). Subsequently, these transformed factor axes are subjected to normalization to yield the final comparison matrix. No statistical error term is available to convey the accuracy of the comparison thus rendered, but the many comparisons of independent factor studies of abilities presented in Ahmavaara's (2) monograph go far to indicate the probable psychological validity of the method.

Results

The rotated factor matrix obtained for the present sample of 156 male school administrators is shown in Table I together with the loadings of the CPI scales on four factors defined in the study of the university student sample by Mitchell and Pierce-Jones (11). The 18 scales of the CPI are listed in Table I in the order shown by the Manual (Gough, 8), and are also grouped into the four classes of measures proposed therein by Gough. In fairness to Gough, it should be understood that he has not proposed that the four classes into which the CPI scales are grouped have anything to do with empirically defined factors. On the other hand, his rational scale groups can be examined in relation to factor analytic findings, and, perhaps, should be so considered. A study of the factor loadings shown in Table I and of related information from the CPI Manual resulted in the following factor descriptions.

Factor I, the most important in this study as well as in that of students, had its highest loadings for CPI scales named Self Control, Good Impression, Tolerance, Sense of Well Being, Achievement via Conformance, and Intellectual Efficiency. Although intuitively there might seem to be some differences between Factor I in this and the earlier study, we have again been disposed to name it "Adjustment by Social Conformity." The results of our transformation analysis, reported below, will, of course, be brought to bear on the question of factor similarity.

Factor II, in the present study of administrators, had its major positive loadings in the same five CPI scales which were highly loaded on Factor II in the analysis for the students, i.e., Dominance, Capacity for Status, Sociability, Social Presence, and Self-Acceptance. The negative loading (-.30) of Self Control and the moderate positive loading (.34) of Psychological Mindedness on Factor II do not alter the earlier proposal that this factor be designated "Social Poise" or perhaps "Extraversion."

Factor III appeared to be defined somewhat differently by the patterns of factor loadings obtained in the two studies. For the male school administrators, Factor III did not load the Responsibility and Femininity scales appreciably as it did for the largely female student sample. However, a negative loading for the Flexibility scale was associated with positive loadings in the Socialization and Communality scales for both samples, although the Flexibility loading appeared

* Footnotes are presented at the end of the article.

higher for the administrators. It should be noted also that the loading of Achievement via Conformance on Factor III was higher for the administrators, while the loadings of Achievement via Independence are both discrepant and different in sign for the two samples; the administrators exhibited a -.23 loading compared with .11 for the students. In naming Factor III for the students, it appeared reasonable, at least provisionally, to call it "Superego Strength." For the administrators, however, the negative loading of Flexibility, the near absence of a loading for Responsibility, and the trend away from Achievement via Independence suggested that the name, "Inflexible Conformity to Conventional Standards" might be more appropriate. The similarity of Factor III for the two samples remains to be evaluated more objectively by transformation analysis.

Factor IV seemed to describe "Capacity for Independent Thought and Action" for the students (11), but with loadings of .45 or higher for Tolerance, Responsibility, and Achievement via Independence—and with negligible loadings in scales named Intellectual Efficiency, Flexibility, and Capacity for Status—this factor, in administrators, may better be considered to represent "Responsibly Independent Non-Authoritarian Attitudes."

It is one thing, of course, to make intuitive comparisons of independent factor analytic studies, and quite another to judge the similarity of factors from more objective evidence such as is provided by Ahmavaara's method already described. The method of transformation analysis was applied to the two factor matrices shown in Table I to obtain the comparison matrix of invariance values shown in Table II. The invariance values ranged from .99 (nearly perfect) invariance for Factor II through .01 to several negative values, which Ahmavaara (2, pp. 133-134) treats as indicating "... that there has been no invariance at all." It would be meaningless to state any average invariance value for the table.

Discussion and Conclusions

There are at least three points around which discussion can properly center and about which tentative conclusions may be drawn. These are: a) the factorial composition of the CPI when applied to city school administrators; b) the configurational invariance of CPI personality factors across camples from different populations as revealed by independent factor analyses performed by identical methods; and c) the validity, in factorial terms, of the four classes into which Gough (8) has ordered the 18 scales of the CPI.

In a descriptive way, the four factors obtained for the school administrators seem psychologically reasonable and relatively devoid of surprises. Social adjustment and extraversion syndromes have been generated by analyses of personality inventory behavior for years (cf.,Tyler, 16). More interesting, perhaps, are Factors III and IV obtained for the administrators. In a sense, they suggest two variable kinds of administrator personality: a) the relatively inflexible follower of the norms of the dominant culture or subculture, and b) the relatively non-authoritarian, responsibly independent leader. It would be informative to have estimates of the loadings of various administrative effectiveness criteria on these personality factors, for both administrative theory and executive selection might profit from such knowledge.

From the standpoint of factor analysis, the four classes of CPI scales suggested by Gough (8) and shown in Table I, fared rather poorly in the present study of school administrators; this was true also in the original factor analytic study using student subjects. Again, however, we would conclude that Gough's Class I, "Measures of Poise, Ascendancy, and Self-Assurance," has been factorially substantiated, although the Sense of Well-Being scale should probably be excluded. Our Factor I, Adjustment by Social Conformity, has substantial loadings on three Class II scales, in two Class III scales, and in one scale each of Classes I and IV. Similarly, our Factors III and IV for male school administrators loaded in CPI scales located in two or more of Gough's rational scale classes. It seems likely that gains in the interpretability of CPI profiles would accrue from reorganizing the CPI scale classes along lines suggested by factor analyses such as the two independent studies which have now been reported.

The issue of configurational invariance from one to another analysis is a thorny one, and the methods presently available for making necessary comparisons are beset with various difficulties (Bargmann, 3). Nonetheless, it is probably better to approach the problem through relatively objective methods such as Ahmavaara's, which is based on an explicit, carefully developed rationale, than to try to solve it in a wholly judgmental way. The results of our transformation analysis shown in Table II support the conclusion that the four factors identified in the student sample were relatively well substantiated in our study of school administrators. However, the transformation analysis was based on only two independent studies, and Ahmavaara's convention (2, pp. 132-133) is to consider factors whose invariance values are thus known only from a single comparison to be of relatively low certainty. Nevertheless, he has accepted invariance values of .50 or higher from a

TABLE I

ROTATED FACTOR MATRICES FOR THE CALIFORNIA PSYCHOLOGICAL INVENTORY: SCHOOL ADMINISTRATORS (A) AND UNIVERSITY STUDENTS (S)

CPI Classes and Scales	I		II		III		IV	
	A	S	A	S	A	S	A	S
C-I. Poise, Ascendancy, Self-Assurance Measures								
1. Dominance	10	18	63	76	16	04	01	-09
2. Capacity for Status	39	21	56	59	-22	04	09	50
3. Sociability	23	27	71	78	-04	04	01	15
4. Social Presence	05	-01	69	62	-20	-23	15	51
5. Self Acceptance	-17	-17	66	77	-10	08	18	05
6. Sense of Well-Being	67	79	07	16	01	05	12	23
C-II. Socialization, Maturity, Responsibility Measures								
7. Responsibility	22	58	22	09	08	44	52	21
8. Socialization	19	43	03	02	63	57	13	-06
9. Self Control	78	92	-30	-19	19	08	08	-01
10. Tolerance	60	67	19	11	-02	15	45	54
11. Good Impression	75	83	06	08	10	-08	06	-06
12. Communality	09	02	-13	06	32	58	-05	-02
C-III. Achievement Potential, Intellect Efficiency Measures								
13. Achieve via Conformance	57	80	29	25	34	23	31	08
14. Achieve via Independence	39	47	05	02	-23	11	54	67
15. Intellectual Efficiency	58	46	30	24	-07	16	18	53
C-IV. Intellectual and Interest Modes Measures								
16. Psychological Mindedness	45	47	34	22	-27	-16	03	32
17. Flexibility	28	-12	01	02	-67	-25	15	56
18. Femininity	-02	-02	-22	-02	-03	45	21	-04

Note: Decimal points omitted before all factor loadings in the table.

single comparison as tending to substantiate a factor (2, p. 133). Following this practice, we have reached the following provisional conclusions.

1. Factor I of the school administrator study is apparently very similar to Factor I identified for the student sample by Mitchell and Pierce-Jones (11).

2. Factor II of the analysis for administrators appears to be virtually identical with Factor II identified for the students.

3. Factor III for students exhibited a relatively strong relationship to both the third and fourth factors found for administrators, an important result that merits further examination here. Factor III in the analysis for the student sample was named "Superego Strength," and appeared to be best viewed as a sensitivity to moral and ethical issues and a conscientious desire to resolve such issues in terms of fundamental values commanding the individual's strong allegiance. The fact that Factors III and IV for administrators were both highly related to the students' Superego Strength dimension suggests, however, that superego strength may be differentiated according to the sources of the person's controlling values. Thus, one variant of superego strength might result where value systems were largely a product of an unthinking, rigid assimilation of conventional standards. Still another sort of superego strength would be expected where values were rooted in reasoned, carefully considered, relatively independent judgments. The moral imperatives would probably be as binding in either case, but their bases of authority would be markedly different. In different individuals, of course, superego strength could center in either one or the other, or in combinations, of these orientations. It is highly interesting to note that these two conceptions of superego strength are in essential accord with those derived independently by Peck (12) from his case studies of adolescents.

4. Factor IV of the study of university students was found to have a moderately high positive relationship to Factor IV for administrators and a moderately high negative relationship to administrators' Factor III. Since Factor IV for the students was identified as "Capacity for Independent Thought and Action" while the administrators' Factor III was interpreted as "Inflexible Conformity to Conventional Standards," the negative relationship between the two factors appeared to validate the designations assigned to each. This also appeared to be true for the high positive relationship between Factor IV for the students and Factor IV for administrators.

5. The pattern of relationships shown by Factors III and IV for the administrators (cf., Table II) supported our earlier inference from the factor matrix for administrators (cf., Table I) that two fairly distinct administrative personality variants can be identified. Thus, the administrator who tended to be characterized by a high degree of "inflexible conformity to conventional standards" would probably be conscientious enough, e.g., about his job, thus showing superego strength of a sort, but he would probably demonstrate relatively little "capacity for independent thought and action." On the other hand, an administrator strongly characterized by "responsibly independent non-authoritarian attitudes" would be expected to show an equal degree of superego strength and a stronger "capacity for independent thought and action." It appears, then, that the factor analytic findings for administrators and the results obtained from the comparison of separate factor studies by transformation analysis have provided evidence supporting the postulation of these interesting administrative personality variants.

In view of the relative novelty of its use, it may be well to consider the apparent general value of transformation analyses of the kind conducted in this research. In the present case, the analysis not only yielded information about the relative variance or invariance of the factor patterns found in two different populations, but it also shed considerable interpretive light on the nature of the factors identified for both populations. It well may be that the non-invariance of some personality factors should not be a cause for regret. Perhaps, instead, factor analytic investigators of personality should undertake systematic cross-population studies more frequently. These studies might in time produce more knowledge about personality organization and its antecedents than would result from the consistent confirmation of the same factors in different populations. Some of Cattell's work appears to have been conducted within this general orientation (cf., Cattell et al., 6, p. 144).

Summary

This investigation of the California Psychological Inventory (CPI) was designed to determine the factorial composition of the inventory's 18 scales for a sample of 156 city school superintendents and to compare the factors so found with those obtained in an earlier factor study of CPI scores for 258 university undergraduate students, most of whom were women. In both studies, four CPI factors were defined by identical procedures of centroid analysis and varimax rotation. Ahmavaara's method of transformation analysis was applied to the two factor matrices to assess the relative configurational invariance of factors across the two different populations sampled. The results of the transformation analysis indicated that two fac-

TABLE II

COMPARISON MATRIX OF INVARIANCE VALUES FROM TRANSFORMATION ANALYSIS OF FACTOR MATRICES FOR UNIVERSITY STUDENTS AND SCHOOL ADMINISTRATORS

		University Students			
	CPI Factors	I Adjust by Conformity	II Social Poise-Extraversion	III Superego Strength	IV Independent Thought-Action
School Administrators	I. Adjust by Conformity	.97	.03	-.09	.24
	II. Social Poise-Extraversion	.01	.99	-.05	.16
	III. Conformity to Conventional Standards	.46	.06	.65	-.58
	IV. Responsible Non-Authoritarian Attitudes	.34	-.11	.71	.60

tors, "Adjustment by Social Conformity" and "Social Poise", were virtually identical in the samples from two different populations. The remaining two factors were less similar, but these variations seemed related to differential modes of assimilation of values and to varying manifestations of independent and flexible attitudes. Further research should clarify the antecedents of variations in personality factors across populations. Additional cross-population studies of personality factors, and their comparison by transformation analysis or other appropriate methods, might yield significant new knowledge about personality organization and structure.

FOOTNOTES

1. We are indebted to Dr. L. J. Carleton, Dean, School of Education, Montana State University, who administered the CPI to the school administrators and made the scores available to the senior author. Dean Carleton's testing program was supported by the W. K. Kellogg Foundation through the Northwest Cooperative Project in Educational Administration to which we also express our appreciation.
2. Our original factor analysis of the CPI made use of data from the Mental Health in Teacher Training Project at the University of Texas, a project supported by a grant from the National Institute of Mental Health. The project provided support for the original analysis and assistance in the present one, and we are grateful to the project and its director, Prof. R. F. Peck.
3. Our choice of method was also governed by the fact that one of the present authors, Dr. F. J. King, then at the University of Texas, had independently developed and verified a computational program for transformation analyses using the IBM Model 650 electronic computer. The present research employed King's program.
4. Only those parts of the two factor matrices which have tests in common are used in the transformation analysis.

REFERENCES

1. Ahmavaara, Y., Transformation Analysis of Factorial Data. Helsinki, Finland: Finnish Academy of Science, 1954.
2. Ahmavaara, Y., On the Unified Factor Theory of Mind. Helsinki, Finland: Finnish Academy of Science, 1957.
3. Bargmann, R., "Yrjo Ahmavaara on the Unified Factor Theory of Mind," Psychometrika, 1960, 25, 105-109.
4. Becker, W. C., "The Matching of Behavior Rating and Questionnaire Personality Factors," Psychological Bulletin, 1960, 57 201-212.
5. Carleton, L. J., Relationships of the Rated Effectiveness of School Administrators to Certain Personality and Personal Background Characteristics. Unpublished doctoral dissertation, University of Oregon, 1956.
6. Cattell, R. B., Stice, A., and Kristy, N., "A First Approximation to Nature-Nurture Ratios for Eleven Primary Personality Factors in Objective Tests," Journal of Abnormal Social Psychology, 1957, 54, 143 -159.
7. Fruchter, B., Introduction to Factor Analysis. Princeton, N. J.: Van Nostrand, 1954.
8. Gough, H. G., Manual for the California Psychological Inventory. Palo Alto, Calif.: Consulting Psychologists Press, 1957.
9. Guilford, J. P., Psychometric Methods. New York: McGraw-Hill, 1954.
10. Kaiser, H. F., "The Varimax Criterion for Analytic Rotation in Factor Analysis," Psychometrika, 1958, 23, 187-200.
11. Mitchell, J. V., Jr. and Pierce-Jones, J., "A Factor Analysis of Gough's California Psychological Inventory," Journal of Consulting Psychology, 1960, 24, 452-456.
12. Peck, R. F., "Family Patterns Correlated with Adolescent Personality Structure," Journal of Abnormal Social Psychology, 1958, 57, 347-350.
13. Thomson, G. H., The Factorial Analysis of Human Ability. Boston: Houghton-Mifflin, 1951.
14. Thurstone, L. L., Multiple-Factor Analysis. Chicago: University of Chicago, 1950.
15. Tucker, L. R. "A Method of Synthesis of Factor Analytic Studies." U. S. A. F. Personnel Research Section: Research Report No. 984, 1951.
16. Tyler, L. E., The Psychology of Human Differences. New York: Appleton-Century-Crofts, 1956.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 1, September 1962)

RELATION OF LORGE-THORNDIKE INTELLIGENCE TEST SCORES OF PUBLIC SCHOOL PUPILS TO THE SOCIO-ECONOMIC STATUS OF THEIR PARENTS

WILLIAM F. ANDERSON
Syracuse University

IN VIEW of the recent publication (1957) of the Lorge-Thorndike Intelligence Test, and in view of its increasing use by school personnel, there is a need to know if the scores on this test distribute themselves along a social class dimension similarly to scores on certain other intelligence tests.

The purpose of this paper is to report on an investigation of the relationship between social class and measures of intellectual functioning derived from the Lorge-Thorndike. The questions answered here are: 1) Do pupils of different social classes obtain mean IQ's which are similar? 2) Do pupils of each of the social classes obtain similar mean verbal and non-verbal IQ's?

Although no attempt is made to provide a comprehensive review of the literature which discusses the relationship between social class and intelligence, it has been noted in numerous articles and studies that persons of the higher social classes obtain mean scores of greater magnitude than persons of the lower social classes (Stroud (8), Shimberg (7), Neff (6), Loevinger (5), Havighurst (1), Herrick (3)). Concerning the relation between social class and the factors of an intelligence test which constitute the total score, Havighurst and Breese (2) found that children of higher social status were significantly superior to those of lower status on each of the Thurstone tests of primary mental abilities. Janke and Havighurst (4) found significant social class differences for each of a series of verbal and non-verbal tests.

The Lorge-Thorndike was administered in the spring of 1960 to all pupils attending the fifth and seventh grades of two central school systems in the Greater Syracuse, New York area. Included in this sample were 319 fifth grade pupils and 279 seventh grade pupils. Both school systems had special class facilities for the educable mentally retarded; these pupils were not included in the sample.

The Lorge-Thorndike is designed to measure intellectual functioning from kindergarten to adulthood. According to the authors, it is a test of abstract intelligence which is defined as the ability to work with ideas and the relationships among ideas.[1]* While the authors recognize that most abstract reasoning involves the manipulation of verbal symbols, a parallel set of non-verbal tests is provided.

The method for ascertaining the social class of the participants involved the utilization of Sim's hierarchy of the social prestige associated with occupations.[2] In developing the SCI, Social Class Identification Scale, Sims obtained judges ratings as to the social class of a list of occupations. These occupations were chosen so as to be fairly representative of all socio-economic levels ranging from unskilled through skilled labor to professional. These occupations were translated into one of seven social class levels.

In this study, the occupation of the father of each pupil was ascertained by using school records and through personal inquiry when the title was ambiguous. Such job titles as machinist and accountant required inquiry into the exact nature of the occupational duties. Occasionally, a machine operator was listed as a machinist, or a bookkeeper as an accountant. Interviews with parents and teachers permitted a correction of all vague occupational titles. Once the occupations were ascertained, the social class of each was determined through use of Table II of the <u>Manual of Directions of the Sims SCI Occupational Rating Scale</u>. When an occupation was not included in the SCI, the prestige values of occupations found in the studies of Hieronymus[3] and Anderson[4] were utilized for determination of the social class.

Only six of the seven social class categories

* Footnotes will be found at the end of the article.

TABLE I

INTELLIGENCE TEST SCORES OF THREE SOCIAL CLASS GROUPS

Group	Verbal IQ		Non-Verbal IQ	
	Mean	S.D.	Mean	S.D.
I (N = 48)	117.8	13.6	113.6	10.2
II (N = 213)	105.9	14.7	104.0	14.9
III (N = 337)	104.3	12.4	102.1	12.2

TABLE II

t VALUES AND LEVELS OF SIGNIFICANCE OF THE SCORES OBTAINED BY THE THREE SOCIAL CLASS GROUPS

Groups	t Value	Level of Significance
I and II (Verbal)	3.81	.001
I and III (Verbal)	5.09	.001
II and III (Verbal)	1.37	.20
I and II (Non-Verbal)	4.16	.001
I and III (Non-Verbal)	6.16	.001
II and III (Non-Verbal)	1.81	.10
I (Verbal and Non-Verbal)	1.69	.10
II (Verbal and Non-Verbal)	1.75	10
III (Verbal and Non-Verbal)	2.32	.05

defined by Sims were utilized. The upper-upper class was inappropriate as no pupils were found in the sample whose fathers held positions appropriate to this level. The remaining six categories were combined into three levels. The highest level, Group I, contained the upper-middle and upper social classes. Group II contained the middle-working and middle social classes, and Group III consisted of the lower-working and working classes. The sample appears as follows:

	Group I (UM, U)	Group II (MW, M)	Group III (LW, W)	Total
Grade 5	25	119	175	319
Grade 7	23	94	162	279
Total	48	213	337	598

Results

The means and standard deviations of verbal and non-verbal IQ's obtained by members of three social class levels are presented in Table I. The mean scores on both factors increased with socio-economic status. Group III had the lowest scores on both the verbal and non-verbal scales, Group II ranked next on both measures, and Group I obtained the highest mean scores. Thus, the scores on the Lorge-Thorndike distribute themselves along a social class dimension similarly to scores on many other intelligence tests. Noteworthy is the fact that for each social class level, the mean verbal IQ exceeded the mean non-verbal IQ. This finding seems inconsistent with the commonly held belief that lower social class children are more handicapped in manipulating verbal symbols than with the management of non-verbal problems.

As a means of determining the statistical significance of the differences between verbal and non-verbal factors within a social class group, and the differences among the social class groups, the t test of significance was applied. This technique was deemed appropriate since the analysis of variance (f test) for the nine comparisons that were to be considered was non-significant at the five percent level.

The levels of significance found in Table II indicate that Group I (upper-middle and upper) obtained significantly higher mean verbal and non-verbal IQ's than the other two groups. However, the mean differences between Groups II and III, on both factors, were not sufficiently large to be statistically significant at the five percent level.

Regarding the discrepancy between mean verbal and non-verbal IQ's within each social class level, the difference was statistically significant at the five percent level for only one of the three groups. For Group III (working and lower-working classes) the mean verbal IQ of 104.3 was significantly higher than the mean non-verbal IQ of 102.1. Although the mean verbal IQ for Group I was 4.2 points higher, this difference was not statistically significant. For Group II, the mean verbal score was 1.5 points higher than the mean non-verbal score. This difference was not statistically significant.

Summary

The purpose of this study was to investigate the relation between scores on the Lorge-Thorndike Intelligence Test and the social class of participating pupils. The results indicate that the Lorge-Thorndike yields scores which are highly related to the social class of the pupil's family. Higher social class status is associated with higher mean IQ's on both verbal and non-verbal factors of the test.

There is no evidence that pupils of lower class status are more effective on non-verbal items than on verbal. In fact, the mean verbal IQ for this group was statistically significantly higher than the mean non-verbal IQ. Whether poorer performance of the lower class pupils is a function of native capacity, or motivation, or lack of opportunity for growth of intellectual potential, or some other factor, is not known. However, these results raise doubts regarding the commonly held belief that lower social class pupils are more effective when dealing with non-verbal problems than they are in verbal reasoning.

In view of these results, it might be desirable to repeat the study utilizing a more detailed classification of social class such as the index of Status Characteristics of Warner, Meeker and Eells. It might also be desirable to include pupils attending the special classes so as to have available the full range of intellectual ability.

FOOTNOTES

1. Lorge, I. and Thorndike, R. L., The Lorge-Thorndike Intelligence Tests, Technical Manual, Houghton-Mifflin Co., 1957.
2. Sims, V. M., Sims SCI Occupational Rating Scale, Manual of Directions, World Book Co., 1952.
3. Hieronymus, A. N., Relationships Between Anxiety for Education and Certain Socio-Economic Variables, Ph.D. Disseration, State University of Iowa, 1948.
4. Anderson, W. F., The Sociology of Teaching I: A Study of Parental Attitudes Toward the Teaching Profession, Ph.D. Dissertation, State University of Iowa, 1952.

REFERENCES

1. Havighurst, R. J., "Culture and the IQ," School Review, 1949, 57, 187-189.
2. Havighurst, R. J., and Breese, F. H., "Relation Between Ability and Social Class in a Midwestern Community: III, Primary Mental Abilities," Journal of Educational Psychology, 1947, 38, 241-247.
3. Herrick, V. E., "What Is Already Known About the Relation of the IQ to Cultural Background?" in K. Eells, A. Davis, et al., Intelligence and Cultural Differences, pp. 10-15 Chicago: University of Chicago Press, 1951.
4. Janke, L. L. and Havighurst, R. J., "Relations Between Ability and Social Status in a Midwestern Community: II. Sixteen-Year-Old Boys and Girls," Journal of Educational Psychology, 1945, 36, 499-509.
5. Loevinger, J., "Intelligence as Related to Socio-Economic Factors," Yearbook, National Society of the Study of Education, 1940, 39, (I), 159-210.
6. Neff, W. S., "Socio-Economic Status and Intelligence: A Critical Survey," Psychological Bulletin, 1938, 35, 727-757.
7. Shimberg, M. E., "An Investigation into the Validity of Norms with Special Reference to Urban and Rural Groups," Archives of Psychology, 1928-29, 16, No. 104.
8. Stroud, J. B., "A Study of the Relation of Intelligence-Test Score of Public-School Children to the Economic Status of Their Parents," Journal of Genetic Psychology, 1928, 35, 105-110.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 1, September 1962)

THE CLASSROOM ISOLATE: AN ADDITIONAL DIMENSION FOR CONSIDERATION IN THE EVALUATION OF A QUALITY EDUCATION PROGRAM

HILARY A. GOLD
Brooklyn College of The City University of New York

A RECENT investigation[1]* has indicated an additional dimension to be considered in the evaluation of a quality education program, namely, the classroom isolate. The isolate is the child who is unchosen by his peers in the classroom. He is, therefore, excluded from the group by the members of the group. At best, he participates in classroom activities from a peripheral position of group membership. When one considers, for example, Jersild's research on belonging[2], Sullivan's recognition of the relationship between lonliness and anxiety[3], and Northway's studies of low sociometric status in classroom situations[4], a study of isolates, and of teachers in those classrooms where isolates are located, is recognizable and warranted.

Accordingly, one of the aims of the research was to attempt to discover whether the ideology of the classroom teacher is related to the relative number of pupils with whom the class as a whole indicate they do not wish to associate. A brief definition of terms used in this paper follows.

Definition of Terms

Ideology. Ideology was defined as "an organization of opinions, attitudes, and values, —a way of thinking about man and society."[5] This investigation was concerned with those ideological patterns which may be indicated on a scale ranging from dominative to socially integrative patterns of thinking.

Dominative Ideology. The term dominative ideology was used to denote that pattern of thinking which predominantly maintains that the customs, attitudes, and values of one's own in-group are a basis for the evaluation of all other groups. Dominative ideology, as is dominative behavior, therefore, is characterized by "a rigidity, or inflexibility of purpose, by an inability or an unwillingness to admit the contribution of another's experience, desires, purposes, or judgment in the determination of goals which concern others."[6] The teacher possessing a dominative ideology was identified by his high scores on a modified form of the F Scale[7]. The child possessing a dominative ideology was identified by his high scores on a children's questionnaire.

Socially Integrative Ideology. The term socially integrative ideology was used to denote that pattern of thinking which recognizes the worth and dignity of all individuals and the groups to which they belong. The teacher possessing a socially integrative ideology was identified by his low scores on the F Scale. The child possessing a socially integrative ideology was identified by his low scores on a children's questionnaire.

Isolate. The term isolate was used to connote the status given by the members of the classroom group, as a whole, to those individual members whom they do not choose to associate with. It in no way indicates that these "unchosen" children are isolated of their own accord.[8]

Instruments

In addition to the F Scale and children's questionnaire referred to above, the children in each classroom group studied completed a sociometric questionnaire. This latter instrument was used to determine the leaders and isolates in each classroom setting. The importance of this approach is that the children are able to express themselves in terms of their own wishes for association; indirectly, therefore, their needs and values may be inferred.

Area of Investigation

The responses of the fifth grade children and teachers of a suburban school district comprised the basic data of investigation. This school dis-

* Footnotes will be found at the end of the article.

trict is composed of seven school buildings, six of which house fifth grade classrooms. A total of 919 children, together with their thirty teachers participated in the study. It is important to observe that this school district has what is generally considered a "good school program." Academic results, as measured by achievement in tests are relatively high. The total district, as evaluated in a Metropolitan School Study Council research, is also superior.

The Data

Isolates. Of the thirty classroom groups forming the basis of research only two contained no children who may be termed isolates. The number of individual children who were identified as isolates in a single classroom unit varied from zero to six. When the important variable of class size is included, the data reveal that the percentage of children in an individual classroom who were identified as isolates, ranged from zero percent to 19 percent. The mean number of isolates for each classroom unit was computed at 2.4 children.

Teacher Ideology. Responses to the F Scale indicated that the 30 classroom teachers studied scored within a range of 1.53 - 5.36, with a mean score of 3.23. In addition, a standard deviation of .94 was obtained. This, it is felt, compares favorably to the range of 1.4 - 5.9, mean of 3.90 and standard deviation of .90 received by Adorno and his colleagues when studying a larger group of 449 individuals.[9]

Teacher Ideology and Isolation. A coefficient of correlation was computed between classroom teacher scores on the F Scale and the percentage of children in each classroom unit who were identified as isolates. This coefficient was identified as .63— statistically significant at the .0005 level. This relationship is further clarified by a division of the ranked teacher scores into quartiles and an examination of pupil isolation within each quartile. This distribution is represented in Figure 1. In interpreting these data, it is important to remember that low scores on the F Scale indicate socially integrative ideology, high scores increasingly dominative ideology.

The data presented in Figure 1 indicate that those teachers whose scores on the F Scale are above the median of the group studied have the greatest percentage of children who are isolates within their classrooms. In examining the percentages of isolates who were located in the classrooms of teachers whose scores on the F Scale fall below the median group studied, an interesting phenomenon is seen to occur: those teachers whose scores fall into the first quartile have a greater percentage of children who are isolates within their classrooms than those of their colleagues whose scores fall into the second quartile. A number of possibilities suggest themselves as related to this phenomenon.

Firstly, it is possible that teachers who fall into the first quartile of the rankings obtained from the administration of the modified form of the F Scale, and who therefore possess socially integrative ideologies, manifest this pattern of thinking in laissez-faire behavior with regard to the guidance of the children in their classrooms.

Secondly, those teachers falling into either extreme of the rankings obtained from the administration of the F Scale may exhibit a rigidity or inflexibility of opinions, attitudes and values. This ideological pattern is, by definition, descriptive of the dominative individual. It is possible, however, that some teachers who attained low scores on the F Scale, and who have been classified as possessing socially integrative ideologies, are socially integrative in a rigid sense. To illustrate: it is possible that these individuals do not permit themselves to see all points of view apparent in a situation, but restrict themselves to only "socially integrative" perspectives.

It should be noted, however, that of the thirty classroom teachers studied, only two did not evidence any children who may be termed isolates in their classrooms. In both of these cases, the classroom teacher scored below the first quartile of the ranked teacher scores resulting from the administration of the F Scale. Thus, the data support the conclusion that the relative number of isolates within the classroom unit is theoretically lowered when the ideology of the classroom teacher is predominantly integrative.

Other Findings

A comparison of the pattern of thinking possessed by the isolates and the pattern of thinking possessed by the other pupils in their classroom groups, as indicated on the children's questionnaire, revealed that the isolates generally possessed more dominative ideologies—or in other words, less socially integrative ideologies— than their peers. Figure 2 represents the percentage of children who are isolates in their respective classrooms, who scored above or below the mean pupil score of their classmates. Thus, 79 percent of the isolates located in this research possess more dominative ideologies than their classmates average; 21 percent possess more socially integrative ideologies than their classmates average.[10] It is important to recognize, however, that the isolates do not differ markedly from the ideological pattern exhibited by the pupils in their classrooms. In no case did the isolates score below or more than two standard deviations above

the mean pupil score of the pupils in their classrooms.

Summary

The data obtained with reference to ideology and isolation within classroom situations appear to indicate that the classroom teacher possessing a dominative ideology fosters isolation. In addition, evidence supportive of Jennings' finding that isolation appears as the opposite extreme to leadership on the "continuum of inter-personal sensitivity between the members of the group and the individual"[11] was found.

It may be assumed that the classroom teacher fosters a psychological climate within the classroom setting. This climate in the case of teachers possessing dominative patterns of thinking appears not conducive to: 1) eliciting the inter-personal contributions of those children possessing relatively less socially integrative patterns of thinking than their peers, and 2) developing the inter-personal relationships which the members of the classroom group are capable of making with one another.

Implications

It has been shown that a great many fifth grade children in the school district studied are not enjoying an optimum learning environment in their classrooms. It has also been shown that the ideology of the classroom teacher is an important variable to be considered in evaluating the total learning situation. Teachers whose ideologies are characterized by high scores on the F Scale were shown to be least effective in the classroom with respect to certain social criteria basic to a quality education program. This study has indicated, therefore, that teachers must have a broad understanding of all they are teaching and encouraging in the classroom learning situation. It is felt that the teacher possessing a dominative pattern of thinking, and most probably manifesting dominative teaching techniques, is actually doing what he thinks is best for the children in his classroom. It is hoped that the data resulting from this study will motivate him, and other teachers, to do research on their own classroom groups. Thus, teachers may discover whether isolates are present in their classrooms, and, if they are, take steps to alleviate this situation.[12]

The problem the isolate presents is indicative of the responsibilities that teachers have to their students and to themselves. They must constantly explore the many strivings, facets, and powers that are part of the on-going classroom situation. The achievement of quality education, in its fullest sense, can only be accomplished by a professional pursuit of the multiple factors present and powerful in the classroom learning environment. Too often, a single standard— such as academic achievement— is judged, presented, and accepted as evidence of a good school situation. Quality education, however, is manifested in the actual lives that children live in the classroom situation—and learn to apply to the larger world outside. Quality education is characterized by children having e-quality of opportunity within the classroom; where children, according to their individual needs, are free to explore, free to experiment, free to make mistakes—and free to think and to learn as individuals. The classroom isolate, therefore, is an additional dimension for consideration in the evaluation of a quality education program.

FOOTNOTES

1. Gold, Hilary A., Ideology, Leadership and Isolation in Classroom Situations, unpublished doctoral dissertation, Teachers College, Columbia University, 1960.
2. Jersild, Arthur T., Child Psychology (New York: Prentice-Hall, Inc., 4th Edition, 1954), p. 257.
3. Sullivan, Harry Stack, The Interpersonal Theory of Psychiatry, Helen Swick Perry and Mary Ladd Gawel, Eds. (New York: W. W. Norton and Compsny, Inc., 1953), p. 262.
4. Northway, Mary L., "Outsiders, a Study of the Personality Patterns of Children Least Acceptable to their Age Mates," Sociometry, Vol. VII, No. 1, 1944, pp. 10-25.
   Northway, Mary L. and Wigdor, Blossom T., "Rorschach Patterns Relative to the Sociometric Status of School Children," Sociometry, Vol. X, No. 2, 1947, pp. 186-199.
5. Adorno, T. W., et al., The Authoritarian Personality (New York: Harper and Brothers, 1950), p. 2.
6. Anderson, Harold H. and Brewer, Helen M., "Dominative and Socially Integrative Behavior of Kindergarten Teachers," Applied Psychology Monographs, No. 6, 1945, p. 9.
7. Adorno, op. cit., pp. 255-257.
8. Adapted from Helen Hall Jennings, Leadership and Isolation, 2nd Edition, (New York: Longmans, Green and Co., 1950), pp. 78-9.
9. Adorno, op. cit., p. 258.
10. This is a reversal of the pattern observed with reference to the relative ideology of the children's classroom leaders and the pupils in their respective groups. See Gold, op. cit., pp. 95-102.
11. Jennings, op. cit., p. 205.
12. Suggestions which teachers using sociometric techniques may use, at their discretion, in

helping pupils who are isolates or outsiders to the group are outlined, for example, in: Mary L. Northway, A Primer of Sociometry (Toronto: University of Toronto Press, 1952) p. 38.

FIGURE 1.

PERCENTAGE OF CHILDREN WHO ARE ISOLATES IN EACH QUARTILE OF THE RANKED TEACHER SCORES ON THE F SCALE

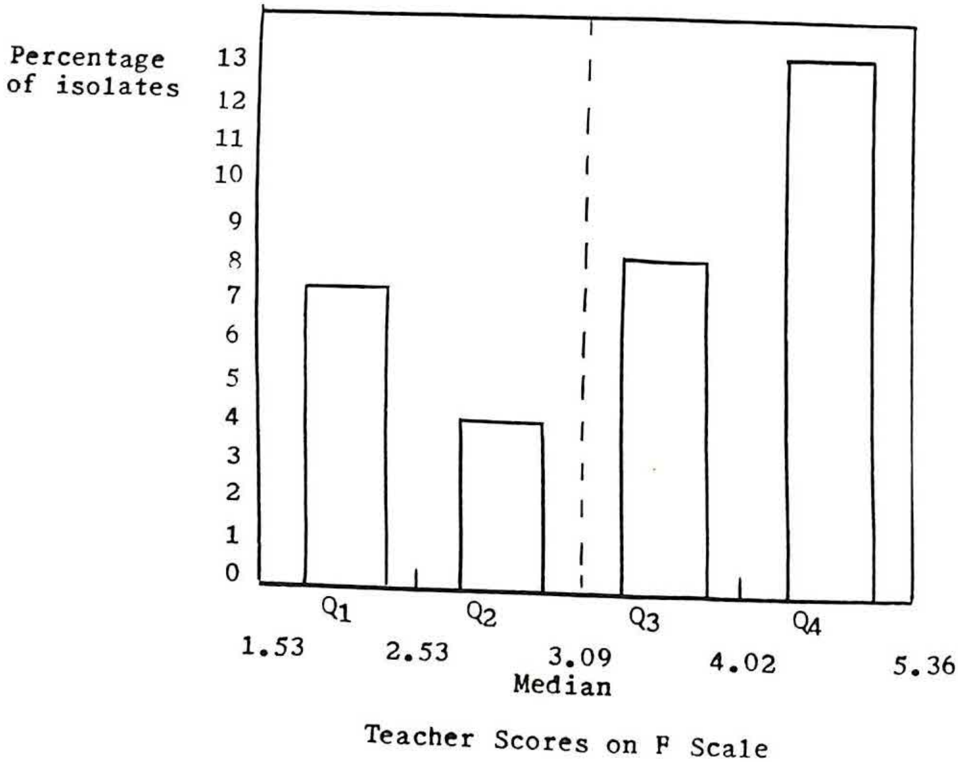
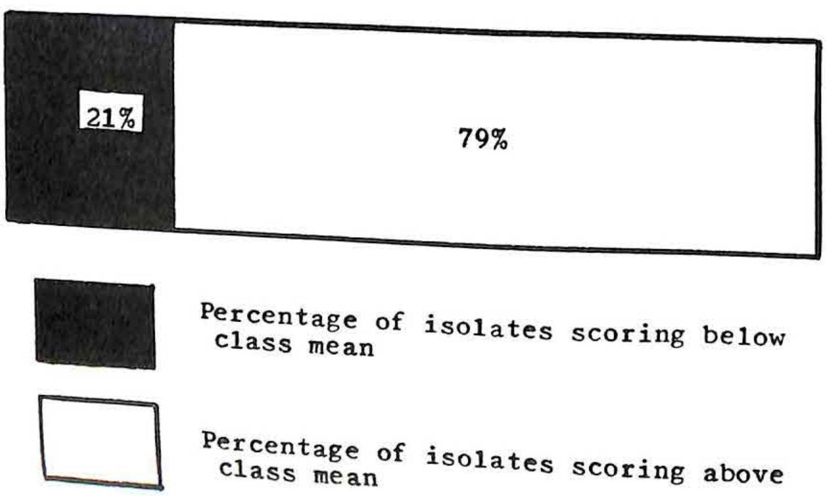


FIGURE 2.

RELATIONSHIP OF ISOLATE SCORES TO MEAN CLASS SCORES



THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 1, September 1962)

A NON-MATHEMATICAL QUANTITATIVE APTITUDE TEST FOR THE GRADUATE LEVEL: THE QED

ROBERT E. STAKE
University of Nebraska

ONE OF the objectives of graduate study is the improvement of competencies for dealing quantitatively with various subject matters. An ability to quantify, to discriminate, and to judge is the mark of the scholar. Our search continues for ways of predicting which graduate students will become the scholars.

The GRE

Currently, most attempts at formal prediction of success in quantification at this advanced level are based on scores from the Graduate Record Examination, Quantitative Aptitude section. The items of this test are of a type common to scholastic aptitude tests for high school students, with emphasis on arithmetic reasoning, algebra, and interpretation of graphs (1, p. 336). It has been suggested that an attempt should be made to devise a type of item especially for this sophisticated clientele, a type which would itself elicit responses indicative of a talent for dealing with complex and subtle quantifications.

The DMRT

Such an attempt was made by Doppelt in the early 1950's. He authored a test designed to evoke an inductive process, with items involving basic quantitative concepts couched in mathematical terminology. This Doppelt Mathematical Reasoning Test is currently available, but seems to be in limited demand in the social sciences, perhaps because it is new or perhaps because of its use of the language of the mathematician. An attempt was made by Stake in 1954 to devise a test of the same item type without reliance on the mathematical language (2). This test was named the Quantitative Evaluative Device, the QED.

The QED

More specifically, the QED was designed to be an instrument to measure certain human abilities that might correlate with eventual sophistication in numerical data treatment and graphic presentation, in statistical inference, in deductive and inductive reasoning, and in definition and categorization of symbols, words, and objects. (Some have thought it strange to call precise use of words a quantitative task. Precise expression is different, of course, from range of vocabulary fluency of expression.) This test, though entitled a quantitative test, was not intended to be a test of mathematical sophistication. An effort was made to de-emphasize the effect of mathematical training even though it was realized that amount of formal mathematical training is somewhat correlated with talent for dealing with quantitative matters. It was felt that an indication of potential for quantitative sophistication would contribute more toward the assessment of entering graduate students than would an indication of present level of accomplishment.

The items of the QED are, like Doppelt's, inductive items. As shown in Figure 1, each item has five choices. In each item there are five possible groups of four choices. One of these groups is a better group than the other four groups. The examinee is instructed to select the group of four choices that is the most integral group and to indicate his selection by marking the choice that does not belong to the group. The correct group can be identified by a quantitative property, such as symmetry, central tendency, or monotonic curvature, but the examinee is not required to know such labels. It has been suggested that the QED would provide more information if the examinee were told to explain or entitle his choice of a group, but the author felt that this would be distracting and time-consuming. It was not seen to be important whether an examinee got the item right because of a strong intuitive hunch or because he recognized the mathematical implication of the category.

Characteristics

The first form of the QED was taken independently and voluntarily by 52 advanced degree recipients who had been enrolled in the department of Educational Psychology and Measurements, Uni-

FIGURE 1.

SAMPLE ITEMS FROM THE INSTRUCTION SHEET FOR THE QED

The directions indicate that one choice does not share a property that the other four have in common. The correct answer is that choice which, when eliminated, leaves the most unified group of four. Correct answers for these items are d, d, e, b, a.

A. (a) 2
(b) 4
(c) 8
(d) 5
(e) 6

B. (a) a bushel
(b) a heaping teaspoon
(c) a handful
(d) a mile
(e) a cubic centemeter

C. (a) Immaculate
(b) Abominable
(c) Worthless
(d) Extraordinary
(e) Usual

D.

(a) (b) (c) (d) (e)

E.

(a) (b) (c) (d) (e)

TABLE I

RANK ORDER CORRELATION BETWEEN QED SCORES AND RANKINGS OF QUANTITATIVE EFFECTIVENESS OF N GRADUATE STUDENTS BY 18 PROFESSORS

Professor	N	Rho	Professor	N	Rho
A	66	.89	J	8	.62
B	13	.86	K	7	.57
C	27	.83	L	8	.52
D	28	.82	M	7	.41
E	9	.77	N	6	.38
F	7	.75	O	6	.26
G	7	.72	P	16	-.04
H	22	.65	Q	7	-.23
I	8	.62	R	7	-.54

versity of Nebraska. Faculty members rated these former students on the effectiveness with which they had handled various quantitative aspects of their programs. The criterion ratings were more highly correlated with the QED scores, $r = .80$, than with scores from the Miller Analogies Test, $r = .48$. (The MAT was not devised, of course, for this criterion.) The Pearson r between MAT and QED scores was .51. The QED odd-even reliability coefficient was found to be .88 on this occasion (2).

The QED was revised three times on the basis of internal consistency item analyses, inspection, and comments of examinees. In one analysis, particular attention was given to items which discriminated between the lowest quarter and the lower middle quarter. It was felt that selection of graduate students in most schools is a matter of discrimination among the poorer and better individuals in the low half of the group applying for admission.

Each of the forms used to date, Forms A, B, C, and D, are revisions of the initial experimental form, not parallel forms. Form D has 65 items, is administered in one hour, and for 925 graduate education majors at the University of Nebraska, has yielded a mean score of 33.9 and a standard deviation of 7.1. The matched half reliability coefficient was .80 for this form and this group.

More on Validity

Coefficients of correlation between statistics grades and the QED scores (Form B) for education majors at the University of Minnesota were not high, .30 and .17, but (with Form C) were higher at the University of Nebraska, .58 and .38, and at the University of Missouri, .83. These groups involved 40, 22, 26, 19, and 18 students, respectively. (The Pearson r between QED and DMRT scores for the Nebraska and Missouri groups averaged around .65. The Pearson r between QED and the MAT for 71 Nebraska graduate students was .44). In the instances where both QED and DMRT were used to predict statistics grades, the two predictors did essentially equally well, the DMRT having a slight edge.

Form C of the QED and the Graduate Record Examination were administered to 459 second semester seniors and graduate students at Long Beach State College in California. The coefficient of correlation between the QED and the quantitative aptitude scores was .59, and between the QED and the verbal aptitude scores was .52. No validation information was obtained at Long Beach.

In a later attempt to validate the QED, 18 social science professors were asked to evaluate a small number of graduate students whom they knew very well. The criterion rankings were to be made as indicated in the following statement.

"Each professor is to base his judgment on how well the graduate student can handle the quantitative aspects of research and advanced study. A student who stands high in this respect is one who is able to work effectively with raw data and with tabular and graphic summaries of data, who appreciates the power and limitation of statistical inference, who reasons well, and who discerns subtle differences and similarities wherever they may exist. He is not necessarily a student who has had formal training in higher mathematics, although it is to be expected that good mathematicians would rate high on these criteria. The evaluation should be based upon the student's performances at the graduate level or post-doctoral level of study. It should not be based upon the results of standardized tests."

The results of previous QED testings, available to over half these professors, were transformed to ranks for each of the 18 groups. Group sizes and Spearman rhos are shown in Table I. It was noted that the median validity coefficient was .62. On this basis, the QED was considered to be valid for use at least by certain professors and with certain groups.

The QED has been administered to over 2500 graduate students and has been used in the screening battery at three independent colleges. Casual reports to the author on examinee reaction suggest that as far as examinees in the social sciences are concerned, the QED is more tolerable than the MAT or the DMRT.

Summary

The Quantitative Evaluative Device has been developed to predict successful pursuit of the graduate college objectives which involve quantitative transactions. The test has been revised in accordance with generally accepted techniques. Evidence of its predictive and concurrent validity has been presented in the preceding paragraphs.

REFERENCES

1. Buros, O. K. Mental Measurements Yearbook. New Brunswick, New Jersey: Gryphon Press (1960).
2. Stake, R. E. Predicting Success in Quantification at the Graduate Level, unpublished masters thesis, University of Nebraska (1955).

The Journal of Experimental Education, Volume 31, Number 2, December 1962

BASIS FOR DECISION

A Composite of Current Institutional Research Methods and Reports for Colleges and Universities

L. J. LINS

Professor and Coordinator of Institutional Studies, the University of Wisconsin, *Editor*

TABLE OF CONTENTS

FOREWORD

"The key to effective administration is the ability of the president and those who work with him to ask the right questions and then find the right answers. But the right answers to the right questions, whether they are specific in relation to a given institution or whether they are more comprehensive, must take into account all the relevant, factual data--the kind of data that only institutional research can provide."*

Through institutional research, whether as a function of a specific office or whether resulting from separate reports of various offices, problems facing institutions of higher learning are brought to light and bases for decisions provided. The need for evaluation may be as diverse as the purposes for which institutions are operated. Thus the conception of what institutional research encompasses may vary from one college or university to another.

In preparing this publication, the editor sought to provide reports some of which might be useful to persons in each institution irrespective of its type, size, or goals. With this in mind, manuscripts were solicited covering a wide range of topics.

This publication is presented as a service to persons interested in and concerned with institutional research. It is hoped that it will, at least in a small way, fill a felt need--that of providing an avenue for exchange of methods and results of research.

The editor is very grateful to the various authors for the manuscripts provided and for their cooperation in making our first publication of this type possible. Credit for the articles is reserved for the authors; should typographical errors appear, these are through my oversight. I appreciate also the efforts of Mrs. Elizabeth Hough, Edna Hasse, Allan Abell, and Robert Rees, of my staff for typing, proofing, and editorial assistance. Special thanks are due the late Professor A.S. Barr and Professor C.A. Schoenfeld for encouragement and offering the avenue for publication.

I respectfully dedicate this monograph to Dr. Barr, in recognition of his many kindnesses to me. His passing represents a great loss to many of us personally as well as to education in general.

L. J. LINS
Editor

* A. J. Brumbaugh. Research Designed to Improve Institutions of Higher Learning (Washington, D. C.: American Council on Education, 1960), p. 2. (Reprinted with permission of the American Council on Education.)

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

INSTITUTIONAL RESEARCH, PLANNING, AND POLITICS

LORING M. THOMPSON
Northeastern University, Boston

COLLEGES AND universities are now at the fountainhead of social change. They have teamed with government and industry to accelerate research in such areas as weapons systems, automation, space travel and communications. The consequences of this research will react through our social milieu and pose new political problems. In addition, they will affect the universities themselves, their social structure, their organization, and their methods of operation.

Proceding more slowly, a movement has begun for universities to study themselves and to plan consciously for changes within themselves. In the past, leadership and changes in universities have come about through strong presidents or influential faculty committees. Additional organizational patterns are now being developed so that universities will be better able to make consciously planned changes.

Institutional Research

One of these new patterns is institutional research. An office of institutional research in a university is expected to make studies about the university itself. Just what features are studied varies from institution to institution. Historically, institutional research may be traced back to self-studies of academic programs by faculties and faculty committees. More recently, enrollments and prospective enrollments have received attention from bureaus of institutional research. With the tapering off of the "veterans' bulge" in the late 1940's, university enrollments reached a temporary valley because of the small number of births during the depression 18 years earlier. However, enrollment projections emphasized that the future would not be like the present, for the rising birth-rates of the 1940's would bring in a "tidal wave" of enrollments. Seen in a broader perspective, this "tidal wave" is actually a cyclical fluctuation about the long range trend for this century. During the time since national records have been kept, enrollments have increased at an average rate of between three and four percent per year.

Projections of enrollments relate to another phase of institutional research, that of studying the capacity of the physical plant and its efficient utilization. Many universities have studied the utilization of classrooms and laboratories during normal working hours and have faced the uncomplimentary conclusion that their efficiency is very low. In some instances, improvements have been made by giving the responsibility for room assignments to the university registrar rather than letting each department or college control designated classrooms solely for its own convenience.

The college students themselves have been a focal point of some bureaus of institutional research, particularly when the director comes from the field of sociology or psychology.

Another focus of institutional research has developed in large state universities. Because legislators tend to make appropriations in terms of enrollments, the institutions have studied the cost of teaching not only the average student but also the freshman, sophomore, junior, and senior student. Cost accounting, taught by the universities for students preparing for business, has been applied to the universities themselves, the motivation being the usefulness of cost accounting data in obtaining state appropriations.

University Physical Planning

Distinct from the development of institutional research, there is another movement in the direction of conscious change in many of our universities today. This is the appointment of planning officers or the establishment of an office of university planning to deal with physical aspects of the campus and its buildings. The university planners draw up campus master plans and become involved in problems of land acquisition and land use and in broad questions of architectural design.

The university planning movement has been spurred by Section 112 of the Urban Renewal Act. Before the passage of this section, universities located in blighted urban areas were in difficult straits. Being tax exempt, they were not included in renewal programs aimed at increasing city income from the property tax. Land they dearly wanted for expansion could be assembled by the city through eminent domain, but only for other users participating in an urban renewal project.

Briefly, Section 112 provides distinctly new ground rules for the relation between an urban university and its host city. It makes it financially feasible for city officials working in co-operation with university planners to include the expansion of an urban campus as part of an urban renewal program. For this purpose, it is practically essential

that a university have an official designated as a planning officer.

Strategic Research and Planning

In the broader sense of the term, university planning involves not only the physical campus but also the long range goals and objectives of the institution, the nature of its educational programs, the kind of students sought and the impact of its programs upon these students, together with the resources (funds) needed to conduct these programs and the means of acquiring these resources (tuition, grants, government appropriations, and fund-raising). Decisions on goals, programs, or fund-raising are strategic decisions of the institution. The decision as to whether or not to seek sponsored research, for example, is a strategic one. It will have a substantial impact upon the type of research on the university campus, the nature of its faculty, and therefore the nature of its students and its future enrollment.

At present, bureaus of institutional research have made few studies related to broader strategic questions such as this. But the university planner must answer these questions, implicitly or explicitly, in his physical plans. If the planning is to be a rational process, then it begins with the determination of long range goals and objectives, followed by identification of the educational methods to be used and the facilities required. Here the planner may readily become involved in pre-planning studies. His area of interest overlaps that of the institutional researcher. Whether the total effort is to be termed institutional research or planning will probably be determined by historical accident. What is important is that the knowledge of modern physical and social science be applied to the analysis of major problems facing the university.

Some Generalizations on Strategic Decisions

At this point, two propositions may be stated concerning the relation between strategic decisions and technical analysis, such as that involved in research, physical planning, operations analysis, or architectural design.

1) The more important, or more strategic the decision, the smaller is the amount of pertinent technical analysis.
2) At an increasing rate, technical analyses are being developed which are pertinent to the more important or more strategic decisions. In other words, advances in research and planning are reducing the extent to which the first proposition is true.

Examples which illustrate these propositions may be drawn from our modern urban society as well as from higher education. Business leaders once made decisions on new products or the location of facilities on the basis of their experience and judgment. Now corporations make extensive technical studies prior to a formal decision. With progress in operations research and management science, it is now becoming possible to make technical analyses of raw materials procurement, production, warehousing, inventory control and distribution--all as a single integrated system. Final decisions made by top management remain the sole responsibility of management, but more and more these decisions are made only after technical studies by persons highly trained in many fields of applied science.

Seen in this perspective, the movements of institutional research and physical planning in universities are part of an important trend within our modern society. According to the propositions stated above, these movements will grow and develop, encompassing subject matter which will make them more and more relevant to the major strategic decisions of universities.

University presidents and faculty committees will not be displaced by institutional research and planning. On the contrary, institutional research and planning can exist only when created and nurtured by presidents and faculties. In the future, one of the skills needed by university leaders will be that of selecting personnel for research and planning, together with providing resources and specifying the boundaries and objectives for the work of the specialists. Once the leaders have the benefit of research and planning work, their decisions will be less limited by their own experience and knowledge. They will be able to examine proposed changes with a better picture of what the consequences will be. Their concept of possible consequences will be derived from careful technical studies, the application of scientific knowledge by competent professional staff.

As this movement develops, we may visualize that future faculty committees will have at their disposal offices of institutional research and planning. These offices will be able to identify possible courses of action which will be realistic, feasible, and in line with the general goals of the institution. On the basis of accumulated scientific knowledge, they will be better able to estimate the consequences of each proposed project or future course of action.

A New Pattern in Politics

It is emphasized that the advent of institutional research and planning will not eliminate internal politics. It will not mean that important decisions will be strictly technical ones rather than political--using the word "political" in a broad sense to refer to the governing of a university. Faculty committees may well have more work to do rather than less because a good office of institutional research will raise questions, make the faculty more aware of evolving changes, and point out new courses of

action which might be considered. To illustrate, an institutional research office might analyze college drop-outs. On the basis of other research studies, it might suggest to the faculty committees new policies and procedures to reduce the number of drop-outs. By itself, the committee could only bring together the opinions of its members on the drop-out problem; with the help of organized institutional research and planning, it can direct that inquiries be made first to pinpoint the problem, then to find solutions.

As institutional research and planning gathers momentum, it should be able to deal not only with problems encountered in the operation of current programs, such as enrollments and schedules, but also with more basic problems of adolescence, motivation, and of learning. It should be able to make long range studies and come to grips with methods for teaching creativity as well as academic excellence. It could make long-term studies of the whole educational process so that it can be better designed to give the younger generation a sense of initiative, responsibility and maturity.

In our era of rapid scientific progress and consequent social change, there are bound to be changes in educational institutions. To make institutional changes, new organizational devices are being developed--both in community politics and in university politics. Recognition is being given to professional specialists who are finding theoretical tools to deal first with minor, tactical decisions, then major or strategic ones. In university politics, this means that offices of institutional planning and research are becoming increasingly important. They will serve university presidents and faculty committees, providing a current, rational basis for leadership in education.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

PROGRAMMING THE UNKNOWN: GUIDELINES FOR THE CONDUCT OF INSTITUTIONAL RESEARCH

PHILIP H. TYRRELL
Rensselaer Polytechnic Institute

UNLIKE MOST other types of programming, such as working out instructional sequences or coding computer inputs, the programming of institutional research is programming the unknown. In establishing guidelines for the conduct of institutional research in the future, one must define concepts, describe assumptions, and formulate strategy. At Rensselaer Polytechnic Institute, programming the unknown has required the development of a rationale that includes all three of these fundamental steps.

Major Concepts

Institutional Research and Educational Research

For purposes of this discussion, institutional research is defined broadly as the methodical study of any problem connected with the operation of the instructional programs of a college or university, together with an attempt to implement a solution to each problem or otherwise effectively utilize the results of such research. This concept includes not only traditional data gathering, processing, and interpretation and the study of operational procedures, but also it includes those activities usually called "educational research," a term whose many acceptable definitions all pre-suppose disciplined, scholarly inquiry into the processes of teaching and learning.

This concept of institutional research is based on a spectrum of investigation so various as to encompass routine survey activity at one end and long, hard, and often lonely projects leading to significant discoveries at the other. In addition, there are implementing-action programs, providing feedback to the investigators. Indeed, there is not a spectrum of institutional research; there are spectra. Collectively, they involve programming the unknown.

The Individual Institution of Higher Education

An office of institutional research must have a frame of reference that defines, philosophically at least, the institution and its role in human affairs.

Most importantly, an institution of higher education should be a center of confluence for many social forces. Into its human and physical networks are swept all known human history as well as the turbulent discoveries of the moment, and from these dynamic networks of interaction flow the hypotheses that help shape tomorrow. Great institutions may have always had this image of themselves, and perhaps this is what Whitehead had in mind when, in Modes of Thought,[1] he said, "The task of a University is the creation of the future, so far as rational thought, and civilized modes of appreciation, can affect the issue. The future is big with every possibility of achievement and of tragedy." Centers of confluence, therefore, point out directions which, in fact, become hypotheses; these, in turn, become objectives, either for institutions or, in the case of very great colleges and universities, for man.

The Office of Institutional Research

The two preceding concepts find combination when one considers an office of institutional research also to be a center of confluence. As such, however, it should be a microcosm that facilitates, through its special kind of research, the activities of the macrocosmic, purposeful, life-liberating assembly of scholars who undertake those individual and collective tasks necessary to "the creation of the future." Not that institutional research is outside the bounds of scholarship; it both serves and is part of the intellectual community. With this image of itself, an office of institutional research may often find an area of great usefulness in the study and implementation of findings concerning an institution's human and physical networks, e.g., in network characteristics which facilitate or impede individual or collective creativity.

Also, for the institution, the image of confluence may help preclude the possibility of existence outside the main stream of history; for the office of institutional research, the image pre-

---

1 Alfred North Whitehead, Modes of Thought (New York: Capricorn Books, G. P. Putnam's Sons, 1958), p. 233.

cludes existence as an impotent bystander in an institutional world of dynamic interaction. Thus, the tasks confronting institutional research include spectra of unknowns, and for this reason, guidelines constantly need to be developed for the future.

Assumptions

Following the construction of a conceptual frame of reference, the effective conduct of institutional research requires the calculation of risk. Ideally, one should be able to estimate the degree of risk involved in at least five major assumptions concerning the individual institution:

1) That the institution is in a continuous process of strengthening the ethical sensitivity of its membership (institutional research results cannot always be accepted with serene objectivity; knotty ethical considerations may arise);
2) That effective communication loops exist or can be created, leading from the institutional research staff to top administration, to the faculty, and back to the research staff sometimes also including such other groups as the community, other institutions, and government agencies, so that feedback is inherent in all communications; and, further, that the basic communication loop be not too much in discrepancy with decision-making channels;
3) That there is or will be at least an average chance that the interests and abilities of present and projected institutional research staff will coincide with the demands imposed by required areas of investigation;
4) That the climate of the institution does or will permit institutional research to become a pervasive "idea," involving all institutional members, each to an appropriate degree;
5) That opportunities can be created for the institutional research staff not only to study defined problems but also to initiate investigations based on "intuitive" recognition of needs in order to discover unrecognized problems, analyze their importance, and pursue selected studies; likewise, that the research staff be so selected that its members will undertake obviously needed studies on a self-assigned basis, motivated by a desire for institutional improvement and the enhancement of systematized institutional knowledge as well as by a desire for personal professional development.

If the preceding assumptions can be made on a low-risk basis, the institution may not even need an office of institutional research. However, if they must be made on a high-risk basis, but with some expectation of a rising level of confidence, one might well proceed to an examination of strategy, which is an art, not a science, although it makes use of scientific methodology.

Elements of Strategy

On a theoretical level, some useful elements of strategy for the conduct of institutional research, in the context we are presupposing, may be outlined as follows:

1) Institutional research should not be merely the asking of and finding answers to questions; it should also be "the criticism of generality by methods derived from the specialism of science," to borrow from Whitehead[2] again as he defines "systematization," a process necessarily inherent in all research, as in philosophy. It may be observed that new ideas, in an institutional setting, often are born as generalities which need constructive criticism before they are susceptible to empirical investigation.
2) The criticism of generalities requires, to use current management science terminology, both a "generalist's" and a "specialist's" approach. The former is largely an administrative function; it provides an arena of speculation that comprises an overall view of the institutional context that the generality impinges upon. The latter is more actionable and more pragmatic; it probes, it verifies a negative or positive value. If positive, it provides transition to empirical inquiry if such is necessary. Institutional research, if boldly conceived, often profits from a generalist management, with specialist support.
3) Although "generality" sometimes has a bad connotation, its value as a mode of thought should not be underestimated. So little is known about the highly valued intuitive process that, until proved otherwise, it seems likely that an intuitive cognition may take the form of generality, as well as of specific insight. In evaluation, perhaps the first question should be, "Is this high-order or low-order generality?" Regardless, the ultimate question is importance, and the determination of importance of institutional ideas requires scrutiny - both by the generalist and by the specialist. Each one's mundane world can become strikingly illumined by the impact of what is really important; it is one responsibility of institutional research to keep the important always before the eyes of the institutional community.
4) The element of importance negates any distinction between what is applied or basic in institutional research. Having importance,

---
[2] *Ibid.*, p. 4

effective research becomes a function of talent, time and money; of these, the management of talent is most critical. Three methods are often necessary: direction, indirection, and non-direction. These are closely related to Assumptions 2 through 5, stated earlier.

Programming

Having a conceptual framework for existing or projected research activities together with assumptions and the elements of a theoretical strategy one asks, "What is programming?" Programming a homogeneous institutional research effort is the manipulation of events so as to attain ever-evolving objectives. If one has thought through his concepts, assumptions, and strategy, programming is not a mysterious process; mainly, it requires dexterity in matching a philosophy of operation with the demands of present and future urgencies. Whatever is important is urgent, although the converse is not always true.

The idea of programming the unknown depends, for its cogency, upon the ever-evolving objectives of the institution. However, the homogeneity of the typical institutional research effort is advantageous. Institutional objectives may be ever-evolving, but when they are pinpointed at any moment in time, programming to facilitate their attainment is a clear-cut, deploying-of-resources process.

End-Product

The desired end-product of creating guidelines is, of course, effective programming, and the effectiveness of institutional research programming can be judged only in terms of the individual institution and at such time as evaluation can be hazarded.

Rensselaer Polytechnic Institute, for example, has been developing the previously outlined rationale of institutional research for almost six years. At this moment, unknowns for tomorrow and for years ahead are being programmed, and re-programmed. At this moment, also, one can recapitulate Rensselaer's programming history in terms of achievement and of error and thereby place a value upon and refine the institutional research operation.

Programming is a continuous process, but, at any time, all of the events which it has patterned are vulnerable to and should be amenable to evaluation. This amenability would not be possible without the articulation of a rationale such as that described previously.

Institutional research, throughout American higher education, is now at a point, where a re-examination of concepts, assumptions, and strategy is necessary. Institutional researchers should take this re-examination upon themselves, lest others do it for them. The opportunity for honest inquiry, humbly pursued, may never have been greater for education; it certainly has never had the resources which exist or have been promised and which must be nurtured, paradoxically, with boldness and with care.

If institutional researchers achieve effective programming, they will facilitate immeasurably the activities of the assembly of scholars whom they serve and of whom they are a part. Concurrently, the total institution as a creative enterprise will be enabled to contemplate without despair a future that "is big with every possibility of achievement and of tragedy."

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

INSTITUTIONAL RESEARCH AND AUTOMATION

EDWARD M. STOUT and IRMA HALFTER
DePaul University

AFTER COMPLETING initial self-studies, oftentimes suggested by accrediting agencies, many institutions of higher education, having a keen curiosity to know themselves, are reexamining their purposes and curricular implementations. Naturally, one of the purposes of such institutional self-study is to help improve the overall effectiveness of colleges and universities. It seems clear that any effective growth or development which culminates in lasting educational improvement must come from within the institution itself. A university must know itself before it can properly relate to its faculty, its students, its alumni, and its public at large.

Such self-examination is frequently concerned with questions such as, what are the goals of the institution and what are the characteristics of its present student body. If, for example, different goals and instructional methods are to be chosen by faculty and administration, would a student body with present characteristics retard the realization of those goals?

As a part of its continuing self-study, and in order to project institutional goals for the future, DePaul University decided upon the development of a research design that would provide as complete information as possible of the intellectual, psychological, and social characteristics of its students and how they learn through the college experience. It should be observed that the research was to be concerned primarily with the learners and then with the relation between certain learners' characteristics and a specific institution's goals.

Objective and Nature of Institutional Research

Before entering upon institutional research, it would be best, then, to define its purposes so it may be related to research as it is traditionally understood. To merit the title "research," institutional research, just as scientific investigation in the natural or social sciences, implies not only empirical evidence and method but also the primary objective of true research: a contribution to theoretical knowledge. This is so even though the inquiry be pursued on a single institutional level. The investigation should produce a design that can be replicated and which can, therefore, be broadened to an inter-institutional inquiry.

Institutional research, therefore, must follow the known traditional steps. As a background to any study, the scientific literature must be evaluated in order to find a rationale for the project. Hypotheses have to be carefully evolved. Strict sampling procedures must be employed, relevant variables must be designated, and a selection of proper statistical measures have to be provided — these to determine the validity of the hypotheses being tested. Hence, the outcomes will be based on more than descriptive data alone. The statistical inquiry of institutional research, as is characteristic of any scientific investigation, should be expected to yield a search for principles with some degree of generality even at the single institutional level. [1*]

Interdisciplinary Investigation

Based upon these canons, institutional research will be distinguished from the investigation of a single scholar or discipline by the fact that it will be interdisciplinary. An institutional examination of student characteristics implies an investigation projected, for example, not only by those expert in the intellectual-academic characteristics of students but also by those concerned with the psychological and social interpretations of higher learning.[2]

Theoretical Contribution

Institutional research should be expected to provide data that has practical usefulness, but institutional research must not be confused with "action" research. Practical policy and theoretical knowledge may make use, for example, of the indexes

* Footnotes will be found at the end of the article

of performances of students, but the primary purpose of the search should be to increase theoretical knowledge, if possible. The contribution to theoretical knowledge is the "main-line" investigation, with practical considerations being second. Both practical ends and theoretical inquiry will benefit by the fact that the basis of the search is institution-wide, that is, data are secured from all the colleges and divisions of the university. Thus, for example, through an objective study of student characteristics, all the divisions of the university become acquainted with the nature and implementation of university educational goals, college by college, and on a university-wide basis. From such an institutionally-based study, objective data become available not only for theoretical inquiry but also for discussion of policy making at various levels.

Research versus Evaluation

The distinction between institutional research and evaluation must also be clarified. Evaluation lies along a fairly well defined continuum such as this: institutional (university) aims and purposes⟶ specific educational goals ⟶ college curricular objectives, course content, and instructional methods⟶ testing the effectiveness of the content and the process for learners⟶ evaluation ⟶ reappraisal on the policy-making level of the total situation. Institutional research on student characteristics will, in addition, consider the college or university goals and purposes in relation to the learner. Significant data for evaluation, reappraisal of the educational objectives, and curricular implementations will result. But the investigation on the nature of the students will continue as an independent, student-oriented investigation of the young learner in a complexus of different learning situations in a supposedly unitary (university) environment. To oversimplify, in institutional research covering the field of students' characteristics, the emphasis throughout will be primarily upon the learner — and a group of learners— who may constitute a sample of a population of learners. In such evaluation, the relationship may be between certain facets of the characteristics of a sample of learners and specific educational goals.

Automation in Institutional Research

For an assessment of the characteristics of students of an educational institution the size of DePaul University to be reasonably and quickly available, it became obvious that recourse would have to be made to automatic data processing. This precluded waiting on a projected five-year schedule of manually gathering and classifying data. Following this decision, two issues immediately emerged as vital: (a) how extensive a body of data was available in various divisions of the university and (b) how quickly could the data be gathered for processing.

Several guidelines, in contrast to the normal processing of cumulative day-to-day statistical information, immediately developed.

(1) Emphasis in data processing had to be concerned with developing the idea of unifying information (not centralizing). This was necessary so that the entire body of information would be available as a service to all offices and divisions of the university for continuing basic research projects.

(2) The processing of data must not involve surrender of control over such evidence but must provide a common source of information. This had the effect of eliminating repeated duplications from the separate offices which formerly had been gathering overlapping data.

(3) It must be emphasized, even at the risk of belaboring what is obvious, that electronic processing does not exist for the mere sake of gathering, counting, and summarizing. The effective way to collect data is to plan with a particular office ways of collecting needed information for the purposes of that office and to incorporate these ideas into the functions of automatic processing equipment and the institutional research goal.

(4) An office whose data are to be processed for the first time should be made acquainted with the machine installation. This must be done before a discussion of the information to be processed is undertaken for then requirements and limitations of the equipment are more likely to be understood and accepted.

(5) It is more important that those charged with data processing be inducted into the nature of educational goals and purposes than educators be acquainted with key punches and tabulators.

(6) Data gathering apart from a research design which tends to hypothesize in advance about the final result is a waste of energy and time. Data should not — indeed must not — be gathered merely because it might prove to be interesting.

(7) Tension may be expected between those who see data processing from the viewpoint of its machine processing and those concerned with the art of research design. Processors usually want extreme uniformity in classification and established categories for long periods of time, whereas researchers insist on flexibility. This is asked for primarily because the gathering of information in a dynamic institution will always be complicated by the diversity of the students, the faculty, and inherent institutional procedures traditionally emplaced from college to college.

(8) Energy must not be diverted, therefore, into fitting the problem to the machine. The problem and the steps necessary for its elaboration must have priority over machine solutions. The

"electronic brains" do not think; men do.

(9) The ideal should be, of course, to keep the machines in operation, but if there is no knowledge to be gained from data, the machines may have to be idle.

(10) Careful distinctions must be made between a mere gathering of facts and an attempt at drawing conclusions and between a gathering of facts and implying value-judgments which result in justifications.

At DePaul, several results accrued from these methods of proceeding. First, the separate offices, through the intermediary of data processing, became aware of the wealth of data available on students' characteristics throughout the university. Second, internally there developed some understanding of each office's needs and requirements. Finally, the institution was made aware of the fact that every office was well informed about current educational research on student characteristics. In truth, it found that these offices had gathered important data but were often so overwhelmed by day-to-day operations that the evaluation of material was by necessity left mouldering.

Benefits of Automation

Although this research design to ascertain student characteristics is on a three-year time table, automation has already yielded the following benefits for additional investigation into student characteristics:

(1) Refinement of hypotheses was possible because automation permitted increasing the number of subsets and enlarging the number of combinations of an increasing number of variables on an inter-college, intra-college, and university basis.

(2) New lines of inquiry could be undertaken immediately to amplify, refute, or extend tentative conclusions.

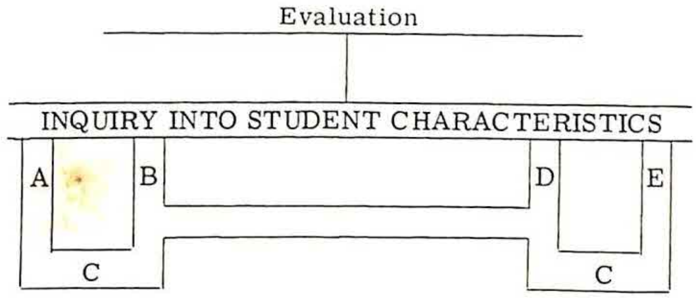
(3) A body of data could be isolated immediately for separate policy decisions without affecting the main line of inquiry.

(4) The offices could evaluate a body of data quickly for its relevance to the purposes for which they were gathered.

(5) Comparative data could affect practical procedures immediately.

(6) Institutional research by serendipity could assume various diagrammatic forms, such as the one at the top of the next column. The legend for this diagram is as follows:

A = University basis of granting honors.
B = High achievers by high schools.
C = Perception of university by high schools.
D = University basis of probation.
E = Low achievers by high schools.



Results

It was possible through the techniques in which automation was used to:

(1) set up a profile of student characteristics (rank in class, scores, and performance) for high-school counselors and college admissions officers; and

(2) reexamine university policies.

Several examples will amplify all six of the above-stated general benefits.

The university grants general admission. That is, it admits to the university and not to a given college, division, or department. The data gathered from the admissions blank, on one item alone — occupational preference — supported the conclusion that registrants had definite academic goals at the time of admission and matriculation. The student interpreted his admission and acted upon the assumption that his admission was differential. Data processing of information gained at the time of admission makes information available for counseling and educational advisement for this differential assessment of student goals.

Since the data seemed to indicate freshmen were entering with a resolved decision about majors, the research design additionally provided elaboration of an investigative instrument to determine what the student's perception of the university was, what the student's perception of himself was (through choice of self-descriptive adjectives), and what the congruence of his attitudes and values was with those of the college.

The important point is that immediate data processing indicated for the current enrollment of freshmen the possibility of a student body not oriented to the exploration of goals. Since occupational preferences usually imply certain value systems, other instruments were designed to establish non-intellective factors as predictors, as criteria, and as contingencies in selection and guidance.[3]

In any study of student characteristics, tuition costs are relevant. Since raising of tuition is almost an annual practice in colleges, electronic processing of socio-economic data yielded such practical information as the size of family from which the matriculant came, the position of the

matriculant in the family, and the financial resources of the student.

In brief, automation is providing at a rapid pace data which are useful for day-to-day decisions, for evaluating academic practices, and for refining the overall research into student characteristics that will advance or retard future educational goals and curricular implementations and will describe adequately the characteristics of learners in a university environment.

FOOTNOTES

1. Walker, Helen M., and Lev, Joseph. Statistical Inference (New York: Holt, Rinehart and Winston, 1953), p. 14.

2. Sanford, Nevitt (Editor). The American College (New York: John Wiley and Sons, Inc., 1962), pp. 5-9.

3. Fishman, J. A. "The Use of Quantitative Techniques to Predict College Success," Admissions Information, I (1957), pp. 49-61.

   Fishman, J. A. "Unsolved Criterion Problems in the Selection of College Students," Harvard Educational Review, XXVIII (1958), pp. 34-39.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

INSTITUTIONAL RESEARCH AT THE UNIVERSITY OF PUERTO RICO

ROSA ESTHER ESCALERA
The University of Puerto Rico

THE UNIVERSITY of Puerto Rico, which is a State Land-Grant Institution, consists of three main campuses located at Rio Piedras, San Juan, and Mayaguez, Puerto Rico. It also has extramural centers in other cities of the island and is about to establish junior colleges in regions where these are most needed. The Agricultural Experiment Station and the Agricultural Extension Service are part of the University, and their functions and services cover the whole island. The University of Puerto Rico is accredited by and has been a member of the Middle States Association of Colleges and Secondary Schools since 1946.

The University ranks eighteenth among the top 50 institutions of the United States according to total enrollment with 21,262 students enrolled the first semester of 1961-62. The University has grown very rapidly; its beginning as a Teachers' Training School was in 1900, but it really was founded by the law enacted on March 12, 1903. It has changed from its original purpose of teacher training to offer a variety of programs of a much broader scope. Enrollment in 1903 was 154; the number of faculty members has increased from approximately 100 in 1903 to 1,664 in 1961. The University of Puerto Rico has grown not only in its student body, number of faculty members, and programs offered, but also it has expanded physically, and continues to improve professionally.

This rapid growth of the institution led the administrative officers to realize the need of statistical studies. The office of the Registrar was the first to start such procedures, but it was on a small scale, dealing particularly with enrollment figures, space utilization, and later on, with data on academic personnel. Some work was also done in regard to entrance examinations and admission requirements.

In 1944, the University of Puerto Rico signed a contract with the International Business Machines Corporation and from that date on, the University Tabulating Division has been renting late model machines. This has given a tremendous impulse to the furtherance of institutional research.

Enrollment figures would be impossible to manage were it not done mechanically. Information obtained from the student on his registration cards is processed in the University Tabulating Division and used in many ways, such as distinguishing regular students from those who do not carry a full-program and preparing enrollment distributions by faculty, sex, and year of classification. The number of courses offered by faculty and department can be determined as well as the number of students per section in each course, student credit hours, faculty teaching load, degrees and certificates conferred, retention rates, etc.

In 1949, the Research Division was created under the supervision of the Registrar, with a staff consisting of its director, one research assistant, two clerks, and 10 IBM operators and supervisors. In 1955, with the establishment of the University Planning Office within the Office of the Chancellor, research was centralized and controlled. The continuous growth of the University and the increasing complexity of its functions brought about the creation of the Office of the Dean of Studies, and the staff of the University Planning Office was separated into those working on academic planning and those dealing with physical planning. Thus, institutional research came to be part of the Office of the Dean of Studies, absorbing academic planning which involves analyses of use of academic space, enrollment projections, and many other statistical projects.

Currently, institutional research at the University of Puerto Rico is centralized in the Office of the Dean of Studies under a research officer whose staff consists of two research assistants and one secretary. Data are collected from the Division of Statistics of the Registrar's Office, from the Office of Academic Personnel, the Office of the Budget, the Offices of the Deans of the various colleges and schools, the Office of the Dean of Students, the Division of Finance, the Division of Extension Programs, and from other offices. Statistical tables and studies are made based on these data which cover all the branches of the University.

Questionnaires from the United States Office of Health, Education, and Welfare, the National Education Association, the National Council on Education, and from institutions of higher learning are answered in this office.

The University of Puerto Rico, being a State institution, depends mainly on legislative assignments for its support and growth. One of the main objectives of the office doing institutional research is to prepare statistical tables and studies to be

used by the Chancellor as a basis for the budgetary propositions submitted yearly for consideration by the Legislature.

Recently, a special study on faculty salaries was made. Questionnaires were prepared and sent to 50 universities and land-grant colleges which were considered similar to the University of Puerto Rico. Replies were received from almost all of these institutions although only 46 were included in the study— a few were returned too late to be considered. The purpose in this study was to compare the salaries paid to the faculty members of the University of Puerto Rico with those of the 46 similar institutions selected. This report is confidential, but it has been distributed to the Chancellor, the Deans, and other administrative personnel of the University of Puerto Rico, for their criticism and recommendations. So far, the responses received are enthusiastic, and it is hoped that this study will be of value to the administration.

Data from graduate schools was used separately from that for the undergraduate level. The comparisons and conclusions attained are considered to be more reliable on the latter level because this sample was larger and more representative than the sample on the graduate level. Mean salaries were used as the basis rather than medians – few institutions indicated their median salaries. Following a comparison of the mean salaries used in this study with the medians published by the National Education Association in February 1962, on a study of faculty salaries of institutions of higher learning in the United States, it was concluded that there is no significant difference between the basis of this study and that of NEA.

With the data obtained from these institutions it was possible to make tables and graphs showing comparisons of the mean salaries of full-time faculty members by rank, and it was possible to reach various conclusions. Names of institutions were coded because of the confidential nature of the study.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

A NEW RE-ACCREDITATION PATTERN BASED ON INSTITUTIONAL RESEARCH: THE SELF-STUDY AT FORDHAM UNIVERSITY

FRANCIS J. DONOHUE
Fordham University

THE RE-ACCREDITATION pattern employed by the various regional accrediting agencies has typically involved a very broad but rather shallow institutional self-study to compile answers to an extensive questionnaire supplied by the agency, an inspectional visit to verify the response to the questionnaire items, and preparation by the visitors of a report on strengths and weaknesses of the institution.

Wide-spread criticism of this typical pattern has concentrated on the "busy-work" nature of the self-study questionnaire items, the strong temptation to camouflage institutional weaknesses rather than to delve into their causes, the failure of the reports of many visiting committees to do more than to underscore strengths or weaknesses which were already obvious in the data supplied to them, the arbitrariness of many of the visiting inspectors, and above all the general futility of the pattern as a stimulus to major institutional improvement.

In an effort to evolve re-accreditation patterns with more positive values, the Commission on Institutions of Higher Education of the Middle States Association of Colleges and Secondary Schools authorized several experiments in institutions of different sizes and types, each designed to try out the feasibility of a particular new pattern proposed for future re-accreditations.

Procedure Pattern

The pattern being tried at Fordham University, as a representative of the "large and complex" type of institution, involves study in depth of five large areas selected by the administration of the University itself as involving serious problems on which the advice of outstanding off-campus consultants might be found useful. The five "problem areas" thus selected were institutional objectives, university organization, faculty, student personnel, and graduate programs.

With a minimal central project coordination staff, consisting only of a full-time Executive Director (who bears the additional title of Assistant to the Academic Vice-President) and one secretary, emphasis has been placed on the work of a central Advisory Committee and of 12 study committees. The Advisory Committee consists of 27 members chosen from the university administration and all eight of its degree-granting units, on five different campuses, and includes scholars currently active as teachers or researchers in 14 different disciplines.

The 12 study committees, ranging in size from six to 13 members each, have a total membership of 102 persons, including a small number of individuals serving on more than one committee. Their areas of study include 1) the functions and responsibilities of the faculty; 2) faculty recruitment, selection and orientation; 3) faculty-University relationships; 4) improvement of communication and mutual understanding; 5) off-campus units of the University; 6) academic re-organization; 7) special academic programs; 8) adequacy of resources for graduate study; 9) material provisions for instruction; 10) residence facilities; 11) non-academic services; and 12) calendar revision.

One notes that none of the study committees are working in the area of objectives. This is because the determination of the objectives of the University was given the fullest time priority, with the Self-Study Advisory Committee itself and the University Academic Advisory Council devoting a series of meetings to formulation of a "Statement of Objectives." This topic is treated as a separate section of this article.

Since the beginning of the Self-Study in June, 1960, the Executive Director has conducted institutional research in such more or less routine areas as faculty load, class size, space utilization, and evaluation of the catalog series. Of greater importance, however, has been the coordination of the work of the study committees and the stimulation and facilitation of research by University faculty and staff members in such fields as student and faculty images of the University, comparison of senior characteristics with the nation-wide findings of the National Opinion Research Center in its study of the career aspirations of college seniors, student retention, and prediction of academic success.

Several of the study committees also have conducted research studies of a questionnaire type. One has done a good deal of primary historical research on the origins of one of the University's units.

Submission of Study Results

The report of each study committee, as it completes either its total assignment or a significant unit thereof, is submitted to the Advisory Committee for extended discussion, along with written comment by the Executive Director. By prior agreement, the study committee reports also are submitted without alteration to the University administration, so that recommendations may be implemented, rejected, or reserved for later submission to the consultants supplied by the Middle States Association.

Regular periodic re-inspections of certain professional units by five specialized accrediting agencies (in the fields of business, chemistry, law, pharmacy, and social service) were held jointly in February 1962 with two Middle States representatives serving as General Chairman and as Vice-Chairman, and the Executive Director of the Self-Study serving as coordinator. This phase of the pattern was also experimental in that the two Middle States representatives conducted no inspections but were responsible for interpretation of the University as a whole to the representatives of the specialized agencies.

The data collected, and the discussions held, up to May 1962 indicate major policy questions on which the University desires outside consultation, in the areas of University organization, determination of the enrollment to be accommodated in future years, and delimitation of the areas in which Fordham University should maintain programs at the master's and/or the doctoral level. Tentative solutions for these policy questions will be prepared by the Self-Study Office and the University administration, and will be submitted with the necessary background papers in February 1963 to a committee of ten consultants selected mutually by the University and the Middle States Association secretariat as being outstandingly competent in matters relating to the areas which the three major policy questions involve. Chairman of the Committee will be Dr. F. Taylor Jones, Executive Secretary of the Middle States Association's Commission on Institutions of Higher Education. After visiting the University and conferring with members of the staff, the consultants will prepare a report to the University which will embody the impressions of their visit and their recommendations to the administration on the specific problems presented to them.

The value to Fordham University itself, with the new Self-Study pattern only two-thirds completed, is already recognized as being much greater than would be that of the typical re-accreditation self-study procedure. The eventual action of the Middle States Commission with regard to changes in the normally required re-accreditation pattern cannot of course be predicted this soon, but the Fordham University experiment is expected to influence at least to some extent the future pattern of re-accreditation procedures.

Faculty Opinion in the Development of a Statement of Objectives

Probably no policy statement issued by any college or university is of greater importance to its faculty and administration— and, in some unfortunate cases, more ignored by both – than is the institution's official Statement of Objectives. If the statement is to be more than verbiage, the faculty must feel that it is their guide to action, a part of their professional life.

Only very broad faculty participation in the drafting of objectives can assure the degree of faculty understanding of institutional objectives, and of integration with them, which will make the objectives a vital force in the operation of the institution. Broad faculty participation has been found to achieve, especially in large and complex institutions, because of mechanical problems in the analysis of faculty opinions and of the synthesis of a document which will express at least the central tendencies of the opinions.

This section of this paper describes a process recently employed at Fordham University to canvas faculty opinion as to the validity and acceptability of a proposed Statement of Objectives, to secure detailed comments and suggestions as to the use or expression of each particular concept included in the Statement, and to determine what changes suggested by individual faculty members were sufficiently close to the central opinion of the faculty as to warrant incorporation into the document.

Inadequacy of Previous Statements of Objectives

The history of Fordham University revealed a series of statements, none intended to be complete in themselves, which over a period of 125 years had attempted to describe Fordham's objectives. Some of these were historical in approach, some philosophical, and some curricular, yet all tended to share three common characteristics: Virtually every one was 1) written by an individual, 2) authorized by some official with the rank of Dean upward, and 3) ignored by the faculty. Up to 1960 there was no official University-wide Statement of Objectives and no general policy on the publication of University objectives in the catalogs of the individual schools and colleges. Some schools and colleges published a statement, others did not, the published statements were in some cases not even consistent with each other. During 1958-1959 and the summer of 1959 a committee of Fordham College faculty members prepared a petition to Phi Beta Kappa for the chartering of a Fordham College chapter of that

Society.* This petition contained a statement of the general objectives of the University, as developed by the committee, but the statement was insufficiently comprehensive to include in its scope the activities of the University's various professional schools.

Lack of an adequate Statement of Objectives was one of the five "problem areas" chosen by the Fordham University administration for study in depth as part of the three-year experiment outlined in the first part of this article. The Self-Study Advisory Committee of 27 members assigned high time priority to the study of objectives. After two meetings devoted largely to the study of the University objectives proposed by the Phi Beta Kappa Committee of 1959, the Advisory Committee rejected them as incomplete because "All Schools of the University were not considered when these Objectives were written....."

Preliminary Statements of Objectives

At this time the Advisory Committee agreed that in any Statement of the Objectives of Fordham University the following concepts must be included: 1) opportunity for students to share in the common human patrimony of letters and science; 2) opportunity for students to acquire and develop the disciplines and modern techniques of the various arts, sciences and professions; 3) integration of religious and secular knowledge and wisdom; 4)liberal arts tradition; 5) advancement of human knowledge by research; and 6) obligation to community (described by some as "Christian maturity").

A "Preliminary Statement on Objectives," based on these factors, was prepared by the Executive Director of the Self-Study, reviewed by the Academic Vice-President, and presented for discussion at a meeting of the University Academic Advisory Council in November 1960 and at a later meeting of the Self-Study Advisory Committee.

Points which particularly aroused discussion at the latter meetings were 1) the relationship of Fordham to the New York metropolitan area, 2) the significance for University objectives of the fact that Fordham is a Catholic and Jesuit University, 3) whether or not specific reference should be made to the University's policies concerning non-Catholic students, and 4) whether or not there should be specific mention of Fordham's "membership in the fraternity of American Universities," which had been emphasized in the Phi Beta Kappa Report but omitted in the "Preliminary Statement on Objectives."

These discussions, and the apparent concensus of the Advisory Committee, were used as the basis of a "Tentative Statement on the Objectives of Fordham University" which was distributed in January 1961 to the 354 full-time members of the staff and faculties of Fordham University who held faculty rank or occupied administrative positions of an academic or quasi-academic nature. With each copy of the "Tentative Statement" (which had been mimeographed in a format in which individual concepts were numbered serially in order to facilitate reference to concepts by number) was enclosed a letter from the Executive Director of the Self-Study explaining briefly the background of the "Tentative Statement" and inviting comments on it; and an 8 1/2 X 13" mimeographed form which the respondents could use to record their opinions.

Faculty Response to the Preliminary Statement

Within approximately two months 198 responses had been received, a return of about 56 percent. The importance of the matter in the minds of members of the faculties is indicated by the attainment of this response without any pressure whatsoever on faculty members to reply. Use of follow-up devices of all sorts was carefully avoided in order to ascertain what percentage of returns could be secured without them, and in order to avoid weakening the consensus by including replies from persons not really interested in the matter.

The distribution of the 198 responses according to the general nature of each, and the unit of the University with which the respondent is affiliated, are shown in the following Table. There were 157 responses generally favorable to adoption of the "Statement" and 24 expressly negative. An additional 17 were mostly negative in their responses to individual concepts.

It is interesting to note in the Table that University-level administrators and members of the Arts and Sciences faculties[1] were very similar in their percentage of generally "favorable" and generally "unfavorable" responses and were more critical of the "Statement" than were any other groups with the exception of the Education faculty. The Education faculty not only had the second highest percentage of response but also the highest percentage of "unfavorable" responses. The two lowest percentages of total response were from the highly professionalized faculties of Law and of Social Service. Those responding from Social Service unanimously gave general approval while the "cagier" lawyers were relatively generous in specific suggestions but apparently reluctant either to approve or disapprove the "Statement" as a whole.

*Fordham College is a liberal arts college for men only, and was one of the original units of the University. The chapter was authorized in 1961, and a charter issued under the designation of "Tau of New York."

[1]Fordham College and the Graduate School of Arts and Sciences, which share one integrated Faculty.

DISTRIBUTION OF 198 RESPONSES RECEIVED TO "TENTATIVE STATEMENT ON THE OBJECTIVES OF FORDHAM UNIVERSITY" BY GENERAL NATURE OF RESPONSE AND BY UNIVERSITY UNIT FROM WHICH RECEIVED

General Nature of Response	Number of Responses, by University Unit*							
	Grad. Sch. Fordham College	School of Bus.	School of Educ.	Coll. of Phar-macy	School of Law	Sch. of Social Serv.	Admin-istra-tors**	Total
Favorable, plus additional favorable comment but no specific suggestions	3	3	6	--	--	2	1	15
Favorable, with neither comment nor suggestions	29	20	8	3	2	3	6	71
Favorable, with suggestions	32	11	14	6	--	4	4	71
Unfavorable, with suggestions	12	1	6	--	--	--	2	21
Unfavorable, with no suggestions	1	--	2	--	--	--	--	3
Neither favorable nor unfavorable, but with suggestions	6	2	3	1	3	--	1	16
Neither favorable nor unfavorable, with no suggestions	1	--	--	--	--	--	--	1
TOTAL	84	37	39	10	5	9	14	198
Percentage of unit's full-time faculty represented in responses***	56.4	68.5	81.3	90.0	35.7	45.0	--	55.9
Percentage of favorable responses	76.2	91.9	71.8	90.0	40.0	100.0	78.6	79.3
Percentage of unfavorable responses	15.5	2.7	20.5	--	--	--	14.3	12.1
Percentage of respondents offering one or more suggestions	59.5	37.8	59.0	70.0	60.0	44.4	50.0	54.5

* The College of Philosophy and Letters (an off-campus residential liberal arts unit), was not originally included in this particular Study. The Graduate School and Fordham College share one integrated faculty.

** University administrators not assigned to any specific School or College. Two administrators from the School of General Studies, which has no full-time faculty of its own, were included in this column.

*** Items on this line are in some cases only approximately correct since some responses from Deans and Assistant Deans are included; the statistics of "full-time faculty members" used had not included administrators.

In all, over 300 specific suggestions for additions, deletions or changes in the "Statement" were made by the 108 respondents (out of the total of 198 respondents) who offered concrete suggestions. These were analyzed by the Self-Study Office and are reported in detail elsewhere.[2]

As illustrative of the suggestions offered by the faculty respondents to the 21 concepts included in the "Tentative Statement," an analysis of the faculty reaction to one concept, "As a member of the contemporary American university fraternity...," is presented.

Two respondents thought this concept should be deleted as having "no real meaning," and two others thought it "superfluous." Another thought it unrelated to the concepts which immediately followed. One respondent thought that it was "too restrictive in that the University's actions seem to be limited to membership in a national professional group in the contemporary scene." One person suggested that it would be desirable to "rephrase or delete" the concept; three others who did not particularly object to the clause as a whole wanted to delete the word "contemporary" (One wrote "contemporary is otiose."). Several objected to the word "fraternity," with two suggesting instead the word "community" – one would write "scholarly community" instead of "university fraternity," and another suggested "scholarly community in search of truth." One respondent suggested that Fordham be described "as a member of the contemporary family of American universities" while another felt that the wording was "too heavy" and suggested that it would be well to "try" the phrase "a representative American university." On the basis of these expressions of faculty opinion, the concept was omitted from the final draft.

The "Tentative Statement" was then re-submitted with a report on the expression of faculty opinion to the Self-Study Advisory Committee, which after discussion and revision recommended its official adoption by the President and the Board of Trustees of the University.

The new "Objectives of Fordham University," as the final document was titled, was officially approved and promulgated in August 1961, approximately one year after the beginning of the Advisory Committee discussions of its content. It has met remarkable acceptance, even by individuals who had strenuously opposed certain of its provisions. This undoubtedly is because the faculty members, having participated in the formulation of the document, find identification relatively easy.

[2] Readers are invited to write for a mimeographed full report of the study, including detailed analysis of the comments of faculty members, to the author at the Self-Study Office, Fordham University, New York 58, New York.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

THE UNIVERSITY OF NEW HAMPSHIRE SELF STUDY

ALLAN A. KUUSISTO
University of New Hampshire

THE UNIVERSITY of New Hampshire self study, 1959-1960, was initiated as a result of faculty prodding and was conducted by a team of six professors aided by off-campus consultants. This study was limited to an analysis of what might roughly be termed the academic or classroom factor of the University's life. It focussed attention on the numbers and kinds of courses taught and the administrative arrangements for teaching, faculty workloads, and time and space utilization.

Background of the Study

The limits of the study were imposed in part by the faculty's interest in what should be studied. The two faculty bodies suggesting the study, the Faculty Council and the Curriculum Committee of the University Senate, were primarily interested in what might happen to the teacher at the University of New Hampshire when the surge of rapidly increasing enrollments strikes the campus in the mid-1960's. They were disturbed by the enrollments being projected for the University by administrators who had studied the admission problem at various times during the previous five years. As a publicly-supported institution, the University of New Hampshire remains obliged to admit all qualified in-state applicants; it could not, therefore, readily impose a ceiling on enrollment. The faculty sponsors were also disturbed by nation-wide reports of an incipient shortage of qualified faculty. They sought a discussion of how the University could maintain or enhance the quality of teaching in the face of drastically altered circumstances.

Organization and Support of the Study

The sponsors suggested to President Eldon L. Johnson that the study be conducted by a faculty team which would be relieved of all other assignments for the duration of the study. This faculty team would collect data, but outside experts would be called in to evaluate the data and make recommendations. The assumption seemed to be that the sensitive nature of the study precluded home-made policy recommendations. The outsiders' objectivity combined with their expertise would presumably engender more support and respect for the completed study.

President Johnson enthusiastically approved the general concept of the study. He won the support of the Trustees for what ultimately became an $18,000 operation supported by the income from the Spaulding Fund which had been given to the University some years earlier for improvement of undergraduate education. The President provided for a year's study by six faculty members whom he would select, three of whom would be on formal half-time assignment to the study; the three others were to serve as advisors and critics without allotment of time.

President Johnson also endorsed liberal underwriting of the cost of consultants and any other expenses, including travel, that the group (to be known as the Committee on Academic Programs and Teaching Methods) needed to accomplish the task. The service of Miss Doris Beane, the Assistant for Institutional Studies in the Office of Student Personnel, was made available to the group in its data collection. The President made no effort to define the scope of the Committee's study. However, he transmitted the letter of the sponsoring groups to the six-man Committee as a general frame of reference.

Qualifications of and Division of Labor Within the Research Committee

The six members were selected as representatives of broad disciplinary areas: social sciences, humanities, engineering, physical sciences, life sciences, and agriculture. President Johnson's appointees were men who had either evinced strong interests in self study or had had experience in conducting such studies. The Chairman of the Committee was also the Chairman of one of the two sponsoring groups. Two others had been participants in a major curriculum study in the College of Agriculture, and a fourth had had experience in curricular studies in the College of Technology.

The Committee spent the major part of its first eight weeks defining its problem and determining divisions of labor. The major areas of study were divided among the members, including the advisory members who volunteered for a full share of the task. One member took primary responsibility for research on time-space utilization; another chose

faculty workloads; the others divided the task of studying curricular problems and administrative arrangements for instruction. At this point, the first consultant was invited to advise on methodology. Dr. John Stecklein, Director of the Bureau of Institutional Research, University of Minnesota, spent nearly two days in almost continuous discussion with the group on methods which would be both manageable and effective in view of the group's amateur status, its rather rigidly-prescribed time limits, and the scattered or unusable nature of existing data.

Time-Space Utilization

The time-space utilization study followed operational procedures listed in the Manual for Studies of Space Utilization in Colleges and Universities by J.D. Russell and J.I. Doi (American Association of Collegiate Registrars and Admissions Officers, 1957), and those used by the Bureau of Institutional Research at the University of Minnesota. The format for ultimate presentation of the material in this section of the report was suggested by the "Utilization of Instructional Space in the State University System of Florida, Fall, 1958" Board of Control, Tallahassee.

In pursuing the time-space study, the Committee member in charge actually measured existing classroom and laboratory space with a tape measure to obtain as accurate data as possible. Class scheduling information was readily available from the administrative offices. The data were then placed on punched cards and suitably coded. The resulting report compared University of New Hampshire use of instructional space and scheduling of classes with that of other institutions which had utilized similar approaches, e.g., the State University of Florida and the University of Nevada. This proved a most valuable guage of institutional efficiency.

Faculty Workload

The main objective of the faculty workload study was to ascertain 1) whether average teaching loads could be increased, perhaps by providing instructors with more sub-professional assistance, and 2) whether non-academic duties take up too much time. The study depended largely upon a fairly detailed questionnaire which was submitted to faculty members who were present on the campus at the time and who had had more than a year of service. Questionnaires were submitted to 230 faculty members; 213 usable returns were obtained. The questionnaires were supplemented by use of data which the Administration had accumulated through annual instructors' reports and through interviews.

Inevitably, the Committee felt that much of the information resulting from the questionnaires was based not upon fact but upon the faculty's estimation of fact. Because of this, the resulting report warned that the compilation and tabulation of the data should be viewed as a beginning rather than as an end to the study of faculty activities. The Committee used the data to propose some ways in which faculty time could be applied more effectively and to suggest a more accurate and equitable method of evaluating faculty loads, but it also expressed its hope that the colleges and departments would use the data in conducting their own self-evaluations of faculty time.

Curriculum and Administrative Arrangements for Teaching

The study of the curriculum and administrative arrangements for teaching involved a number of devices. Moreover, in attempting to assay the amount of unnecessary duplication and proliferation of courses, the quality of existing programs, the role of large versus small classes, teaching machines, television, library facilities, year-round operations, et al., the Committee immediately found itself in an interesting philosophical quandry.

If the Committee was to collect data on these matters and prescribe alternative courses for the future, there seemed to be a need for a definition of the University's underlying objectives. The question proved sufficiently fundamental to compel the entire group to devote at least a month to an effort to define the objectives and to relate programs to these objectives. The multiple interests of a Land Grant University made the problem tortuous, yet the consensus which the Committee reached on this matter, following considerable discussion with faculty and administrators, went a long way in providing order and cohesion to the subsequent research and evaluation. Immediately, the consensus paved the way toward the determination of guide lines for maintaining the University's objectives and programs in a sound relationship.

The methods utilized in determining the state of the curriculum included analyses of time and room schedules, instructors' grade sheets, and other data on file. Information was gathered for the preceding academic year and an earlier reference year. This material was coded and punched on cards.

Much of the qualitative assessment of the curriculum, as well as reactions to varying administrative arrangements for teaching, was based on interviews and questionnaires. All the Committee members conducted depth interviews of department chairmen and of a random sample of different faculty ranks within departments. Over 100 faculty were interviewed. Questionnaires were sent to a random sample of 10 percent of the student body to ascertain views on strengths and weaknesses in existing programs. The Committee was aided in the evaluation of these questionnaires by selected student leaders.

Evaluation of Data

The Committee departed from its original mandate in one significant respect. Following the collection of data, it decided that it should assume the primary responsibility for determining what the data meant for the University's academic future. Originally, emphasis had been placed on the importance of the outside experts' objectivity in making recommendations. This, however, had to be weighed against the Committee's intimate acquaintance with the special nature of the University of New Hampshire. The Committee's decision to offer its own recommendations was reached after a unanimous endorsement of this procedure by the consultants brought to the campus.

Outside Analysts

The task of evaluation involved utilization of several highly regarded analysts brought to campus for a day or two toward the end of the year. These analysts were selected for their obvious interest in the kind of problems with which the Committee was wrestling. They included Dr. Earl J. McGrath, former Commissioner of Education of the United States Government, who was and is involved in research and writing on higher education at Teachers College, Columbia University; Dr. Philip Coombs, then Secretary of the Fund for the Advancement of Education; Dr. C. R. Carpenter, Director of Institutional Research, Pennsylvania State University; and Dr. Paul L. Dressel, Director of Institutional Research, Michigan State University. A "windfall" consultant, who had been brought to the University for other purposes, was Dr. W. H. Cowley, Professor of Higher Education, Stanford University.

The analysts were used in two ways. They were asked the meaning of the data, and they were interrogated about their experiences and opinions about self studies and the future of higher education generally. The Committee agreed that the week or so given to these concentrated sessions with well-informed visitors was invaluable. The analysts provided assurance to the Committee regarding the respectability of its efforts, well-reasoned direction-markers for reaching policy recommendations, and authoritative support for the over-all study.

Related Conferences at Other Institutions

In addition to analysts' visits, Committee members travelled to nearby institutions to interview faculty and administration regarding educational policies which attracted interest. Illustrative of such visits were those to Dartmouth College to study the three-quarter calendar; to Colgate University to observe the core course curriculum; and to the Connecticut Valley of Massachusetts, to discuss the efforts of the University of Massachusetts, Amherst, Mount Holyoke, and Smith Colleges at establishing a merged venture in higher education.

The Results

The resultant study, complete with charts, graphs, and appendices, encompassed 117 printed pages (double column, 8-1/2 × 11 inches). It included 53 recommendations for change in University policy, ranging from some rather minor matters, such as altering the physical education requirements and changing the nature of the Library Committee, to far-reaching innovations, including the prescription of a University-wide requirement of a four-year independent reading program and a recommendation to departments to prepare a master plan for all aspects of their development for the next decade. The report, entitled "The University of New Hampshire and the Future," was formally presented to the faculty in January, 1961, at an assembly convened by President Johnson for the purpose of giving the Committee members a chance to express their views about the study and to entertain questions about it.

The Committee's assignment had been to gather data and to offer a springboard for discussion, leading to action. Although implementation depended on faculty and administrative initiative, the Committee sought to pinpoint responsibility for action along with its recommendations wherever it seemed relevant to do so. In the months following circulation of the study, Committee members found themselves involved in continuing consultation with administrative and faculty groups regarding implementation of the recommendations. This experience led Committee members to the conclusion that self studies might well give more attention to the problem of what happens after the study is published.

Conclusions

If any general conclusion can be reached about the University of New Hampshire self study it is that a serious and methodical ad hoc faculty venture given both financial and moral support by the administration can serve as an invaluable springboard toward a type of internal ferment conducive to general planning of an institution's future. Once data are brought to light about its dark corners, an institution will not readily be able to settle back into a complacent status quo. A valuable by-product of the University of New Hampshire study has been a recognition of the importance of continuous, soundly-conceived institutional research.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

AN APPROACH TO INSTITUTIONAL COST ANALYSIS*

ROBERT E. HUBBARD
Wayne State University

LATE IN THE fall of 1960, Wayne State University began a campus-wide cost study. This was not the University's first attempt to study its costs of operation. Indeed, every organization makes a kind of cost study, even if somewhat unsophisticated, each time it prepares a budget or issues a financial report. Wayne State University departments and colleges had for some time been making informal probes of their expenditures. Even a couple of all-University attempts had been made to study costs before 1960. Few persons, however, placed much confidence in the results of these early cost studies, and the results were seldom, if ever, cited in support of a budget request or in explaining costs of University operation.

Different circumstances surrounded the 1960 study. For one thing, only a few months before, the State Legislature had enacted into law a requirement that all public universities in Michigan must analyze and report their operating costs each year by student level. Various interpretations have been attached to the term "student level," but it is now quite generally accepted to mean the following five divisions of students: freshman and sophomore, junior and senior, masters, doctors, and graduate-professional students.

In retrospect, this enactment produced two very real benefits for Wayne:

1. It caused the University to devote as many resources as could be spared, both personnel and money, to undertake the necessary cost study.
2. Because the study suggested a level of complexity that cost studies had not previously demanded at Wayne, the assistance of the University's Computing and Data Processing Center was enlisted.

Because adequate resources were available and because extensive use was made of Computer Center facilities, the current cost study has achieved a measure of success which Wayne's earlier cost studies did not enjoy. And because few compromises were made in favor of cheaper or less detailed procedures, a level of sophistication and accuracy has been attained in this study which the cost studies of other institutions seldom attain.

Planning the Study

The first step in planning the study was the creation of an advisory committee. Some 10 persons associated with fiscal and statistical offices in the University constituted the committee, which met weekly for several months to establish the broad study framework. It was recognized at the outset that this study would succeed only if active involvement of many persons in the University was obtained, for an all-inclusive study of costs was bound to affect virtually the entire University in one way or another. Success in obtaining University acceptance of the study was due in large measure to the rather lengthy time taken to discuss the study with many individuals and to adjust it as much as possible, in light of the concerns of these persons. A description follows later in this report of how certain specialized information was collected from certain instructional units (over and beyond the common data collected from all) and how this special information was processed according to the wishes of the units requesting its collection.

The advisory committee agreed that the initial focus of the study should be on the direct costs of instruction, specifically the salaries of teaching personnel and related expenses of the instructional units of the University. There were three reasons for this. First, more than half of Wayne's costs are under the budget classification of "Instruction and Departmental Research," which represents the salaries, supplies, and equipment costs of schools and colleges of the University. Second, the actual costs of teaching courses, i. e., the salaries paid to faculty, constitute the principal expense in this classification. Third, since it is possible to identify the levels of students in the courses, this seemed to be both a convenient and a logical starting point for assigning costs back to student levels.

After careful study of alternate ways of apportioning faculty salaries back to the courses taught, a procedure was adopted which made each member of the academic staff responsible for dividing his University service between teaching and other kinds

*Based upon an address delivered at the Seventh College and University Machine Records Conference held at Wayne State University, Detroit, Michigan, May 1, 1962.

of service such as research and public service. Some universities use arbitrary formulas for dividing the service of faculty among the different functions. But all studies must somehow make an allocation of faculty service, which is really an allocation of the time spent in different service areas, for this is widely accepted as the first step in apportioning faculty salaries to different courses and then to different levels of students in the courses.

In reality, there are three basic steps in this important phase of virtually all cost studies: 1) determining faculty time or energy spent for different services; 2) converting each time or energy element to a percent in order to reflect the portion given to each course and to each other service; 3) multiplying the percent of time given to each course by the appropriate individual's salary in order to assign a dollar value to each course.

The decision at Wayne to require the faculty member's judgment of the division of his time is the basic distinguishing characteristic of the instructional cost phase of its institutional cost study. There was some apprehension, of course, that asking faculty members to divide their University service into components might not meet with the approval of certain faculty. The reporting of suspiciously high percentages of time to "filling out questionnaires" by a couple of faculty members suggests that the concerns were not unfounded.

Faculty members, it was assumed, would become increasingly accurate in their judgments of percent of time spent in different service areas as they make judgments over several terms. Limited evidence suggests that this assumption was sound.

Once the decision was made to rely on faculty judgment in dividing time, development of a form, initiation of data collection, and planning for the analysis of data could begin. It was at this point that Dr. Walter Hoffman, director of Wayne's Computing Center, was first contacted to explore how the Computing Center might aid in the project. Following this meeting with Dr. Hoffman, Mrs. Maggie Goodwin began the task of programming the study. This first program yielded some three and a half million figures spread over 3500 tab pages, all for one semester and for only one phase of the institutional cost study.

Collecting Information from Faculty

The form developed to collect information from faculty has now been used four times. Minor changes have been made since the first use, but the basic approach remains unchanged. The form provides for two kinds of information: 1) the percent of time the faculty member devotes to each of the classes he teaches; and 2) the percent of time he spends in other service, including administration, research, public service, etc.

Every academic person paid from State funds and assigned to an instructional unit (school or college) completes the form. This includes deans, graduate teaching assistants, and part-time faculty as well as the regular professorial and other instructional staff.

The percentages entered on each form must total 100. That is, a faculty member devoting 40 hours a week to University service is asked to consider the 40 hours as constituting 100 percent of his time. A faculty member who devotes 60 hours a week to University service must use 60 as his base, and for him 60 hours represents 100 percent of his time. One of the immediate effects of this decision is that it is possible to increase a percent in one category only if there is a counter-balancing decrease in another. If there is any tendency to overstate the percent of time devoted to one service, such as research, this can be done only if another percent, such as that reported for class instruction, is reduced. One faculty member reported that the only real problem he faced in completing the form was keeping his total percent down to 100.

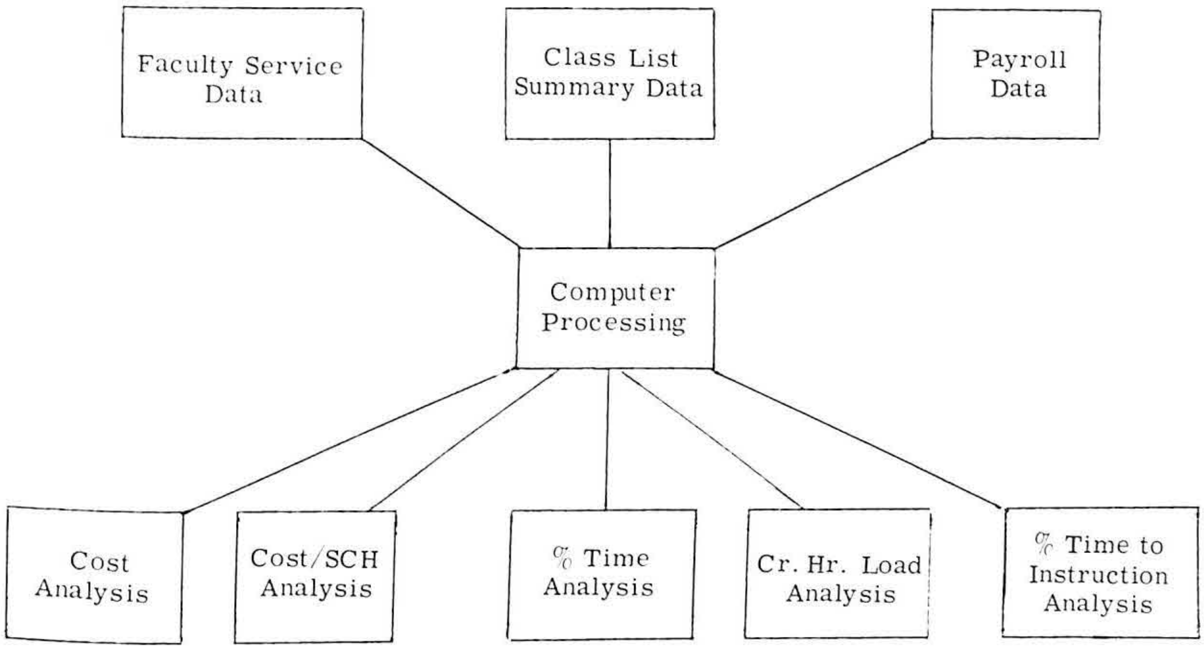
Rationale for Allocation of Instructional Costs

The analysis of instructional costs proceeds as follows:

1. The percent of time an individual reported for a certain class section is multiplied by the salary he received for the semester. The resulting dollar amount is then assigned to that section. If two or more faculty members participate in the instruction of a class, each reports a percent of time for it. The dollar amount assigned to the section is then the sum of the appropriate dollar amounts from the salary of each faculty member participating in the teaching of the section.

2. Once the dollar value for the section is determined, a reallocation of this amount to student levels has to be made. This is done on the basis of the proportionate student credit hours in the class at each of the student levels cited previously. For example, assuming that a faculty member reports 10 percent of his time for a certain class and that his salary is $5000 for the semester, 10 percent of $5000, or $500, is the amount chargeable to the class. If there are 25 students enrolled, each registered for four credit hours, then the class has 100 student credit hours (25×4). If 15 of the students (60 student credit hours) are freshmen or sophomores, then 60/100 of $500, or $300, is charged to the freshman-sophomore level. If the other 10 students are juniors and seniors (40 student credit hours), then 40/100 of $500, or $200, is assigned to the junior-senior level.

3. This basic calculation is repeated for every section reported on each of the faculty service forms; then the dollar values are totalled by student level for each department, for each college, and finally for the entire University. The total dollars so derived are divided by the student credit hour total at each level to yield a cost per student credit hour.



The task of making these thousands of calculations is left to the computer. The input information to the computer, based upon the enrollment and assignments at the end of the second week of classes, is as follows: 1) data from faculty service forms reflecting service to the University for the semester; 2) class list summary cards containing basic student credit hour data by level for all sections taught in the University; and 3) faculty payroll masters from which term salaries can be projected.

The diagram above shows the input and output for the instructional cost analysis.

Output—Cost Analysis

From this program certain costs are obtained based on salaries, fringe benefits, and salaries plus fringe benefits for each teaching classification by department, college, and for the University. To accomplish this, the salary of each faculty member is first assigned to the class sections taught and to other service areas, according to the percentages reported on the form. Teaching costs for each section are then distributed to student level on the basis of the student credit hour distribution, by level, in that section. These teaching costs by level and the other faculty service costs are then summed by teaching classification for each department, for each college, and ultimately for the whole University.

Output—Cost Per Student Credit Hour and Cost Per Student Registration Analysis

In this program, the teaching costs derived from the Cost Analysis are compared to the student credit hours produced by the respective department, college, and the University to yield costs per student credit hour by level. Two different sets of student credit hour figures are used. One is the total student credit hours from all sections reported by faculty. The other is the total of all student credit hours, even those produced by central University personnel and others not assigned to the instructional unit budget. The procedure is repeated for the cost per student registration analysis, where student registrations are defined as the number of students registered in the class sections.

Output—Percent of Time Analysis

This program yields a frequency distribution by faculty head count and by full-time equivalent, showing the number of faculty devoting certain percentages of time to instruction, to administration, to research, and to the other service areas shown on the form, for each teaching classification, by department, college, and for the University. Figures are reported for intervals of 10 percent (from 0% to 100%) but with zero percent and 100 percent shown independently. This is one of the extra kinds of information which the study yields. The output is not directly related to the cost study, but it is useful to different segments of the University in showing how the energies of faculty are distributed over the different areas of service.

Output—Credit Hour Load and Contact Hour Load Analysis

Frequent questions are asked about teaching loads in different departments and for the University as a whole. Measures of load which are typically requested are credit hour loads and contact

hour loads. This program provides a frequency distribution of both credit hour and contact hour loads by teaching classification for each department, college, and for the University. Average teaching loads are also obtained.

Output—Percent of Time Assigned to Instructional Units

This is another special run not related directly to the cost data. Its basic purpose is to provide data for the State Budget Director on the number of full-time equivalent faculty, broken down by teaching classification and by the percent of time assigned to instructional units.

Some Fringe Benefits of the Cost Analysis

This study marked the beginning of a new era in institutional research for Wayne State University. In previous studies, sufficient personnel or time had not been available to utilize the data collected as fully as was desirable, and potentially useful studies had to be eliminated. A whole new dimension was added with this study. It was no longer a question of what had to be omitted. It was, rather, a series of continuing inquiries by our programmer as to what further studies might be desirable.

In the cost study, the immediate concern is simply a complete picture of the way costs should be assigned to specific student levels for the total University. But since the source data begin with classes and individual faculty members, it is possible to make intermediate summaries by department and college before striking the all-University figures. Some day the State Legislature may want costs broken down by department as well as by student level. In addition, this information can be helpful immediately for internal analysis, both by individual departments and by the central University administration. Therefore, all the programs are set up to yield analyses by department, wherever possible.

Frequent inquiries have also been made about differences between the teaching classifications, both in terms of costs and teaching loads. All analyses, where possible, are therefore made in terms of teaching classification.

Furthermore, each of the instructional units is offered the opportunity to collect and have analyzed information beyond that collected in common from all units. The College of Liberal Arts, for example, wanted data on total clock hours of service by faculty. (All units reported contact hours and credit hours of teaching, but only this unit wanted total clock hours as well.) A column titled "For College Use" was added to the form which the College of Liberal Arts used to report clock hours.

Later, the College of Education decided it would be helpful if the "service units" which that college uses to describe faculty service were analyzed. Faculty in that college used the column "For College Use" to report service units, and the Office for Institutional Research arranged for the appropriate analyses.

These "fringe benefits" resulting from the decision to enlist automatic data-processing equipment are emphasized because it seems advisable for each institution to explore such possibilities whenever a detailed study of costs is undertaken. Much of the data on teaching loads either must or can be easily collected at the same time. Not only is a great deal more information made available, but it is also felt that the results from Wayne State University's study are more valid because the greater interest of more people, faculty and administration alike, in the results tends to lead to greater care in completing forms and otherwise providing necessary information.

Allocation of Non-Instructional Costs

All this represents just one phase, albeit the largest phase, of a study of institutional costs. The present program does not yet include a procedure for allocating the costs of faculty service devoted to research, administration, public service, etc., over the student levels. Bases for such allocations are being developed, and when firm decisions are reached, the program will be modified to provide for these.

Neither does the program provide for the allocation of costs other than salaries of academic personnel over the five student levels. These include costs for supplies, equipment, maintaining the physical plant, and the operation of non-instructional units, all of which must ultimately be provided for in the construction of an all-inclusive cost "model." A rationale for the allocation of these costs to student levels is also being developed and will be incorporated in the program.

The Library

During the spring semester of 1962, and each of the past two terms, a sample week was selected as a basis for study of the library with the intent of projecting the week's costs to the full semester. During this sample week, each student taking out a book indicated what his class level was and for what course (hence for what college) he was intending to use the book. Each of the 10,000 or more "circulation slips" was then analyzed to determine what proportion of circulation use was assignable to each college and to each student level in each college. The proportionate use at each level in each college served as the basis for spreading the total circulation costs to these units and levels.

Reference costs for the library could also be identified separately. To assign these to college and level, information was collected during the sample week on the length of time it took to answer each

question asked by each student, identified by level and by course (hence college) for which the question was asked. This system is complex and very time-consuming for it has not been possible, at the present time, to mechanize this phase of the study.

However, the need for information of this kind led to exploring again the use of "charge-a-plates" as student identification cards. This was not the first time the library and other units investigated the possibility of a charge-a-plate identification system. The decision has now been made to initiate a charge-a-plate system during the fall quarter, 1962 (Wayne moves from the semester to the quarter plan in the fall of 1962), thus making the library phase of the cost study as well as the analysis of such areas as the Division of Student Services much less challenging, both in cost and time.

To illustrate how far-out thinking on the potentials of the charge-a-plate system has progressed, the possibility of turnstiles in the library has been contemplated so that the student's identification number from his charge-a-plate could be recorded when he entered and left the library. The time each student spends in the library could be determined by matching input and output turnstile data, and the length of time that students from a given class level from a given college spend in the library could be determined. Since the time spent by students is assumed to bear some relationship to operation and maintenance costs in the library, a basis would be established for assigning library operation and maintenance costs to college and to student level.

Conclusions and Implications for Future Studies

The degree of detail presently being pursued is, of course, much more than is necessary on a continuing basis. However, such great detail in the study seems necessary at least once, although shorter and otherwise improved methods will be developed as the cost study is repeated in succeeding terms. It appears that any university that decides to commit its resources to a cost study must go through a tremendous amount of detailed study which can later be shortened in the light of experience. The important thing is to do it thoroughly the first time so that the proper decisions for adopting abbreviated procedures can be made.

One possible weakness of the cost analysis study is the premise that faculty members are able to make judgments as to how they spend their time. During the past academic year, however, all the Michigan public universities and colleges agreed to adopt, on an experimental basis, the basic procedure Wayne has been following, in order to comply with the legislative enactment cited earlier in this paper.

Clearly this procedure is not a panacea for the sometimes agonizing problem of necessary university costs, but a procedure has been developed which works for this institution and which has elements that may have applicability to other institutions. Even more than the immediate results obtained, there is a long-range benefit which results from the decision to plunge so completely into a study of this complexity. This is an awareness that many of the studies of faculty and students, heretofore rejected because they demanded so much personnel time and appeared to be just too involved and expensive, are really able to be attacked through the facilities of the Computing Center. Some giant strides forward in this area of institutional research, based upon the experience so far with Computing Center personnel and machines, are expected in the next few years.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

UNIT COSTS OF INSTRUCTION IN HIGHER EDUCATION*

D. GORDON TYNDALL and GRANT A. BARNES
University of California

LARGE AMOUNTS of time and energy are being expended on many university campuses in America today in an effort to develop something called "unit costs." In this paper, some of the methodological issues underlying these efforts will be explored, and the methods being used in this area by one university will be described briefly.

It is customary and perfectly appropriate that anyone writing on this subject should first genuflect before that immense project known in the trade as the Cal-Big Ten study. The California and Western Conference Cost and Statistical Study[1] involved a tremendous investment in time and money in a pioneering effort to obtain comparable data from the participating institutions. The Study indicated that it was possible for a group of diverse educational institutions to cooperate in developing comparable cost data, and the participants have made use of the data for a variety of purposes. Unfortunately, the study was not continued, either individually or as a cooperative inter-institutional venture. Perhaps it came too early or the need for the data was not fully appreciated, for funds were not made available to permit their continued collection and processing. Why then is a resurgence of interest in studies of unit costs found even among many of those who did not participate in the earlier study?

The Purposes of Unit Cost Studies

At one level the reasons are obvious. Tremendous growth in higher educational facilities is inevitable in the next decade. It is understandable that legislators, alumni, trustees, and the public should ask what it will cost and hope to get the answer by multiplying the number of expected students by some cost-per-student figure. Administrators respond with magical formulae which attempt to allocate the joint costs of inseparable functions. These efforts do result in some figure which will satisfy the inquiring legislator or the man-on-the-street; beyond that they are virtually worthless.

Nevertheless, there is a pressing need for objective measures and standards in budgeting, planning, and day-to-day management. The sheer magnitude of these tasks, at least in large institutions, demands reliable cost information. Here the pressure is not from a tax-shy public or a legislature bent on trimming budgets, but from those who are facing the difficult decisions: faculty members serving on planning committees, budget officers, and other school administrators; technical staff in the state departments of finance or education; and a few legislators who have made education their special province. What is sought is a basis for control of the tremendously expensive growth that must take place; the goal is to achieve maximum results from limited financial resources.

To speak of control over higher education is to arouse fears. especially among the faculty. It must be emphasized that this discussion concerns only the development of tools of control, without specifying who is to use them. It matters not, for present purposes, whether control is exercised by committees of the faculty, deans, presidents or chancellors, or some agency of state government;[2] the problem rather is how to provide the "controller" with tools to make his control informed and effective. This requires mechanisms enabling the controller 1) to know when and where action is required, and 2) to know what action is appropriate. The first of these has been called "attention-directing" tools; the second, "problem-solving" tools.[3]

Attention-directing mechanisms are required because the controller is unable to have first-hand information concerning all the operations which he is required to control.[4] But directing attention to those areas where attention is needed is no simple task. Even the best methods devised by industry are far from perfect. One frequently finds that the mechanism serves to draw attention to inadequacies in the raw data or in the methods employed rather than to the realities of the situation. Another problem is that some standard must be established for purposes of comparison with the actual performance. This standard may be derived from data for

*Footnotes will be found at end of the article.

the recent past (perhaps an average or a smoothed trend), or it may be based on the experience of other comparable operations; less often, the standard is "engineered," representing what ought to be. However, a standard cannot take account of all the factors that may be relevant in specific situations; in certain cases, a departure from the "standard" may be necessary for the efficient operation of the system.

Some aspects of the operations of a college are similar to those of industrial enterprises, e. g., what is known generally as "Maintenance and Operation." In these areas, it is clear that unit costs will be useful in forming and applying standards. As in industry, if variances exceed the specified norm, attention is directed to the proper control points for possible action.

The process is obviously more complex when one moves to the academic functions, where lines of authority are blurred, the data elusive, and the faculty sensitive to interference. This has led many institutions to abandon attempts to measure and assess the unit costs of their instructional programs. Yet, financial data are sometimes indispensable if rational decisions affecting teaching and research are to be made. Thus, despite the difficulties in devising accurate measures and acceptable standards, it has seemed worthwhile to make serious efforts to develop unit costs in the instructional area.

The Nature of Unit Costs of Instruction

What, then, are unit costs of instruction? They are the cost per some specified unit of providing instruction: for a whole institution, for some school or college, for some discipline or field of instruction, for an individual department, for a single course or even for a particular section of a course. Costs for the larger units may also be broken down by the level of instruction (e. g., instruction in upper division courses in the school of education). In studies of faculty workload, student-faculty ratios, unit costs, and for projections of enrollment the use of the student credit hour (SCH) as the unit of "output" is almost universal in American higher education; indeed, there is a remarkable consistency in the way this concept is applied. We would argue that the SCH is not the best measure in studies of faculty workload or student-faculty ratios; however, since it uniformly measures student achievement,[5] centering upon values attributed by faculty to certain bodies of knowledge and degrees of skill, SCH is the best available index of the "output" of the teaching function.

One must then ask what costs should be identified with instruction, i. e., what input-costs should be related to the output unit, the student credit hour. The most important single determinant of instructional cost is clearly the salary of the teacher with whom the student has direct contact. A few other costs may be attributed to a particular class, e. g., the wages of "readers" used by the instructor in grading of exams, laboratory materials, etc. Also, it may be possible to allocate other costs at the departmental or school level to instruction in particular courses, e. g., mimeographing of course outlines, problems, exams, etc. Fortunately most of these other costs are small compared to salaries and can safely be ignored for present purposes. In the Cal-Big Ten study, an elaborate effort was made to allocate all departmental expenditures on some basis to individual courses. The value of this effort is questionable.

Attempts are also being made in some universities to make detailed allocations of other cost elements, such as library, maintenance and operation of plant, and even general administration. Where legislators or others demand this, the institution may have no choice but to comply; but for purposes of attention-directing or problem-solving, the more complete figure is usually less useful than one which eliminates indirect costs.

Unit Costs and Distribution of Faculty Time

Attention is directed now to two major questions which arise once the cost analysis has been narrowed to departments of instruction:

1. How much of the staff member's time (and salary) should be allocated to teaching?
2. How should that part of the staff member's time (and salary) which has been allocated to teaching be split among the several teaching assignments which he may have?

Some would argue that all staff time budgeted to a department of instruction should be allocated to teaching. This would solve the first problem very simply, and for certain institutions this may be appropriate. However, in some schools, e. g., agriculture, research undertakings are commonly budgeted to a department of instruction, with some professional staff engaged completely or almost exclusively in research, and with some faculty appointments split between teaching and research titles. In these situations, a degree of separation of the teaching and research functions has been recognized as a matter of policy; to attribute all costs to teaching would be unwarranted in these instances.

The question is more difficult where the research is that which has become generally known as "departmental." This is the research which every regular faculty member of a major American university expects as a part of his assignment and which he is expected to carry on in addition to his regular teaching duties. Here the usual argument for the inseparability of teaching and research is qualitative: effective teaching, at least at advanced levels, is best accomplished by one who is himself assaulting the ramparts of knowledge; teaching and

research contribute to each other in countless ways; the university is by definition the place where teaching and research are joined.[6] But the fact that they are inseparable in this sense does not mean that two distinct outputs, teaching and research, cannot be specified and that inputs of time and cost due primarily (if not exclusively) to either teaching or to research cannot be determined with reasonable accuracy. This has importance beyond the mechanics of unit cost studies because there will be some limit, varying according to individuals, where the teaching load is so great that creative research becomes impossible and the effectiveness of teaching and research will suffer. If this situation becomes dominant, the qualities that define a university will be lost. Below this limit—and it is here that a practical problem for unit cost studies arises—the proportion between teaching and research can vary widely as a consequence of the number and type of courses which the instructor is required to teach, the assistance he gets, and the amount of "preparation" which he gives to his teaching.

In the Cal-Big Ten study, faculty workload was computed on a percentage basis, and this was used to allocate a specific percentage of the faculty member's time to teaching. This approach has obvious merits, but given the wide range in the number of working hours per week reported by faculty members, it has one major disadvantage that can be illustrated best by a single example: if two faculty members teach separate sections of a single course, each having the same salary and spending nine hours each week in contact with students and in "preparation" for class (including the grading of exams and course assignments, office hours, etc.), but one states that he spends 25 percent of his (36 hour) work week on teaching, whereas the other spends 16-2/3 percent of his (54 hour) work week on teaching, quite different amounts will be charged to instruction in the two cases by the percentage-of-time approach. This seems clearly unreasonable. To say that teaching costs more in the case of the man with the 36 hour work week <u>because</u> he does less research would indeed be strange; the same cost should be charged in each case.

One obvious method of achieving the same cost is 1) to specify some arbitrary but acceptable official work week, 2) to assign costs by taking the teaching time (including preparation) as a fraction of the work week, and 3) to apply that fraction to the hourly salary rate. Then the second of the two problems—allocating teaching time and salary between several teaching assignments—would be solved automatically, providing information was available from each member of the faculty about contact and preparation time for each of his courses. To date, it has seemed impractical to secure this data, at least without slowing drastically the gathering of data and without arousing the ill will of the faculty. We have, therefore, experimented with a "formula" approach which automatically assigns a given number of hours of preparation to each hour of "contact": 1) two hours of preparation for each hour of contact in lecture and seminar type instruction, and 2) one hour of preparation for each hour of contact in laboratory, quiz, discussion or activity type instruction.[7]

With the above approach, both the size of classes and the units earned by students have been shunted aside. Class size may be an important element in determining the time required of an instructor, but at most large institutions this tends to be offset by the assignment of auxiliary staff—readers and graduate student teaching assistants—to large classes. The time and cost of auxiliary staff must also be applied to the proper course and department. Units earned by students (i. e., student credit hours, the unit value of the course multiplied by the number of students) does not belong on the <u>input</u> side of the analysis, for it is a measure of student achievement, not a measure of faculty time or effort. It would, therefore, be misleading to gauge faculty workload by student credit hours taught. It would also be misleading to use the student contact hour (faculty time spent in class or in contact with individual students) as a measure of faculty workload, because of the widely differing amounts of time spent in "preparation" per contact hour.

The formula specified above was based on a detailed study of faculty time distribution in which 10 percent of the total number of persons bearing teaching titles in each discipline were randomly sampled by mail questionnaire. Respondents were requested to estimate the actual number of hours per week spent during one semester in the usual categories of teaching, research, administration, public service, etc. Teaching was separated by <u>level of course</u> (lower division, upper division, graduate), and <u>type</u> (lecture, seminar, laboratory, quiz-discussion-activity, and individual tutorial); further breakdowns were provided for graduate tutorials and guidance of graduate research. For each class taught (individual students in tutorials), instructors were requested to show separately both contact hours and preparation hours.

The response revealed a surprising consistency in the distribution of teaching time between preparation and contact. The major hypothesis was confirmed: that the ratio of preparation time to contact time varies according to the <u>type</u> of instruction (lecture, seminar, lab, etc.) but is relatively constant among the several disciplines. It was also found that course assignment policies within departments generally correspond to the results of the time study, e. g., fewer unit credits per faculty but more hours of contact in biology departments where the bulk of teaching is in laboratory courses, and vice versa in departments where lecture instruction prevails. When preparation time is applied by formula to individual cases, the teaching work week appears to be quite equitably adjusted among the various disciplines and within particular departments.

Of course, any formula based upon an average of many dissimilar situations is subject to doubt, rigorous sampling notwithstanding. Yet, even if it were practical to get current estimates from all the faculty as to actual preparation time, it would seem better to use the formula approach since this removes certain unwanted factors, e. g., the effect of the extra preparation required when a course is originated, revised, or taught for the first time. Although the effect of atypical situations would not be crucial for any large unit, it would distort the results for a particular course or even a small department.

Data Processing

The faculty time survey thus yields a set of ratios which may be applied consistently to each course offering in order to determine the time consumed in preparing for and teaching the course. Conversion of time to cost is adaptable to electronic data processing. In fact, since the study at the University of California embraces nearly the whole of the University's regular instructional program,[8] amounting to approximately 12,000 individual courses or sections of courses, manual processing would be virtually impossible. As it is, EDP does not do the whole job but, with greater refinement of reporting and programming, it is hoped that manual processing of the data will be eliminated eventually.

Reporting is accomplished with an 11 × 16 inch, 22 column machine accounting form. The information for each course or section is spread horizontally across the form, with entries for course number, section number, type of instruction (lecture, seminar, laboratory, etc.), unit value, scheduled meeting time and place, instructor name, and enrollment by level of student (in seven categories ranging from lower division to post-doctoral). Several other columns are available to accommodate special situations, e. g., hours of instruction for faculty and student, where these differ from the scheduled hours. Certain of this information is now recorded in the registrar offices by hand-copying, but two of the five campuses being studied are able to accomplish most of their reporting directly from IBM class cards with the aid of IBM 407 machines, and it is expected that further automation will be possible within the next year.

The rest of the processing is done centrally at the state-wide administrative office. Here the hundreds of schedules of classes submitted by campuses are edited for conformity to EDP programming, keypunched, and fed to IBM 1401 machines. There are two distinct phases of machine processing: one for the production of space utilization statistics; the other for faculty workload and cost figures.

Workload and cost information is derived from three basic calculations: "weekly teaching hours" for faculty time, "weekly student hours" for student time, and "semester hourly rate" for salary cost.

Weekly Teaching Hours (WTH) includes both contact hours—the scheduled meeting time of the class—and preparation time, supplied by formula according to the type of instruction (lecture, lab, quiz, etc.). Thus, at a preparation time to class time ratio of 2:1, a lecture class meeting three hours per week would carry an allowance of six hours of preparation time, or WTH of nine hours.

Weekly Student Hours (WSH) is simply the product of hours per week of class time and the enrollment of the class. A class of 30 students meeting three hours per week equals 90 WSH.

Semester Hourly Rate is computed individually for each member of the faculty who is teaching. On the basis of two semesters per year, six months of salary is applied to a 15 week semester of 40 hours per week.[9] From this is derived a figure representing the salary cost for one hour of work per week for one semester. (The hypothetical 40 hour week may be said to represent the time "purchased" by the University, though in fact most faculty members work much longer hours.)

These measures are computed and made available by the IBM 1401 for each course and each section meeting of virtually all the departments of instruction in the University. Further, totals of WSH and WTH may be accumulated for departments or groups of departments, schools, colleges, or campuses. They are combined in a number of ways to produce the following measures:

1. Weekly teaching hours per full-time equivalent (FTE) staff. This is the ratio of teaching hours per budgeted staff member. It is a better measure of faculty workload than either contact hours or student credit hours per FTE because it a) allows consistently for the differences in preparation time required for the various types of instruction, and b) avoids the distortion of taking unit value (a "reward" to students) as an element of faculty workload.

2. Weekly student hours per FTE staff. This is a measure of time students actually spend in contact with faculty, weighted by the number of students and compared to the number of staff budgeted for any given department, school, college, etc. It may be compared to the teaching hours per FTE figure, above, as a partial measure of workload; more importantly, it is a better measure of the student-faculty ratio than the old FTE student per FTE staff.

3. Aggregate costs for each class or section, each course, department, college, campus, etc. In its basic form, this is merely the product of WTH and semester hourly rate. This may be extended to include costs other than salary by taking the same proportion of other budgeted expenses as the salary costs bear to total budgeted salary, and pro-rating these to individual classes within a department according to the number of student hours taught in each class.

4. Unit costs. The cost per unit of output is

finally reached when student credit hours are divided into the gross cost obtained above. This may be computed for any one course, department, college, etc.

With appropriate programming, these figures may be broken down either by level of instruction (lower division, upper division, graduate), or by level of student (lower division, upper division, professional graduates, masters candidates, doctoral candidates, etc.).

The Uses of Unit Cost Data

The information described in the preceding section will be made available to appropriate officers of the state-wide administration and, for each campus, to the several Chief Campus Officers. It will permit these officers to obtain a concrete and objective picture of the unit costs of instruction in the areas under their control, in as much detail as they are likely to want or need. But, as has been emphasized above, if these data are to be useful for attention-directing purposes, they must be compared with some standard. In this field it may never be possible to establish an "engineered," objective standard. The present plan is to attempt to develop standards by means of 1) comparison with similar courses (departments, schools) on the same campus or on other campuses of the same institution; 2) comparison with similar departments in other institutions; and 3) comparison with historical data for the same department.

During the early years, this program will necessarily be limited primarily to intra-institutional analysis, analysis of the first type, but continuing and increasing efforts will be made to exchange data with other institutions which have similar programs; as a body of data is gradually accumulated, increasing attention will be given to historical analysis.

It is impossible in the present article to describe the many problem-solving uses to which this data will be put. Extensive use of it is already being made in the budget process, in long-range financial planning of new and emerging campuses, and in the development of program costs, especially in the area of graduate studies. Although what is being done represents only a beginning, it is felt that unit cost studies will become of increasing importance as the problems facing institutions of higher education grow in importance and complexity.

FOOTNOTES

* The authors wish to recognize their indebtedness to other members of the unit and particularly to Mr. Wesley F. Hall, now on the staff of the Budget Office, University of California.

1. A summary report of the findings was published in 1960. The detailed results were confidential and made available only to the several cooperating institutions: University of California, Indiana University, State University of Iowa, Michigan State University, University of Minnesota, Pennsylvania State University, Purdue University, Vanderbilt University, Wabash College, and the University of Washington.
2. For the record, it probably is desirable to note that we do recognize how crucial it is for the future of higher education in America that control be exercised responsibly; among other things, "responsible" in this context means responsiveness to the needs and wishes of each of the groups named, as well as to the public interest. For these reasons, faculty, administration, and legislature should each exercise some control, and in fact, each does exercise some control, though the distribution of control varies widely among the several states. See, e.g., John J. Corson, Governance of Colleges and Universities (New York: McGraw Hill, 1960); cf., Lyman A. Glenny, Autonomy of Public Colleges (New York: McGraw Hill, 1959).
3. For a detailed discussion of this question, see Herbert A. Simon, Harold Guetzkow, George Kozmetsky and Gordon Tyndall, Centralization vs. Decentralization in Organizing the Controller's Department (New York: Controllership Foundation, Inc., 1954).
4. See R. G. Brown and M. J. Draper, "Editing Financial Data for Management," The Controller, XXX, No. 4 (April 1952), pp. 151-57.
5. It cannot be said that the assignment of unit credits, particularly to graduate work, is always equitable from the student's point of view. But this is a minor flaw, and it is expected that the faculty, as they realize the importance of consistency in the assignment of unit credits, will cooperate to correct the situation.
6. To be consistent, however, research accomplished in special institutes and bureaus is equally applicable to teaching, and this introduces further complications in deciding which costs to attribute to teaching.
7. Special methods are employed for crediting preparation time to individual (tutorial) instruction, field courses, supervision of student teachers, etc. In each case, the formula is based on what is known about the particular situation.
8. Only the medical schools and University Extension are excluded. Separate studies are being made of costs in these areas.
9. Most faculty appointments are for nine months, but salaries are paid monthly the year round. Therefore, six months of salary may be attributed to each of the semesters, which are actually in session 15 weeks each. Most of the exceptional cases are handled routinely in machine processing, a few individually.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

FORMS FOR GATHERING DATA ON TEACHING LOADS

CLIFFORD L. CONSTANCE
University of Oregon

IN HIS WORK, the institutional researcher finds there is a need for a complete cross-section measuring instrument to give him concurrent and consistent data on enrollments, teaching loads, space utilization, budgets and costs, and other items. Where computers are used in student registration it is desirable to relate other data to enrollments and to introduce faculty, budget, and time and space information as needed for particular statistical products. This brief paper illustrates how faculty loads can be gathered as a basic statistic for integration with other related data.

The University of Oregon for a long period has used a blank on which individual faculty members record their many instruction-related and non-teaching tasks as well as their direct teaching loads. When used for research, the separate reports raised some questions such as: are two professors reporting the same students for a seminar in which they cooperate; is the graduate assistant conducting a particular laboratory section or is it the same one the professor reports for himself? For this reason, the department or school recapitulation form was devised and added to the regular routine. It may be noted that direct recopying from the individual forms to the summary form is kept to a minimum.

Many of the implications of these forms are quite obvious but may be mentioned. From the individual form, budget and full-time-equivalencies are related to departmental costs and to teaching loads and ratios. Clock hours spent with specific courses are the base for allocating staff full-time equivalents to courses and divisions and for computing student-teacher ratios. Other clock hours permit measuring the load of student advising and of committee work. On the departmental form the total course enrollments on the left of the heavy line are related to course enrollment and student-credit-hour tabulations and studies. Data on the right of the line describe teaching conditions, hence are used for class size, hourly distribution, space utilization, and countless other investigations.

UNIVERSITY OF OREGON, EUGENE

SUMMARY OF INSTRUCTIONAL SERVICE   Fall Term, 19 61

Dept. or School Economics   Signed by Dept. Head or Dean ...

This report is the official recapitulation of this term's instructional work, correlating with and amplifying individual Staff Mem-
ber Reports. It should follow the order of the current Time Schedule and show every course and section listed there or officially
added. (Blank sheets may be used for second and following pages, with the same columns.) This report must be compiled and signed
by the proper department head or dean, and filed (two copies) in the President's Office not later than the fifth week of each term.

LIST EACH COURSE ONCE.

(For variable or arranged credits an approximate average should be shown. Take enrollments as of the end of the second week.)

LIST EACH SECTION OR INSTRUCTIONAL GROUP AS ACTUALLY TAUGHT

(Data should be spread so that no line involves more than one section, more than one room, or more than one instructor; this facilitates valid analyses of space use, faculty load, etc. Indicate any course which is double numbered, and any section which meets jointly with another.)

Dept. and Course No.	Credits per Term	Total Enrollment	Sec. Code	No. Students	Hour and Days of Meeting	Classroom or Laboratory	Name of Instructor
Ec 201	3	485	1A	56	8 MWF	133 Comm	K...
			2A	43	9 MWF	133 Comm	B...
			2B	192	9 TuTh	3 Fen	M...
			3A	43	10 MWF	134 Comm	T...
			4A	49	11 MWF	134 Comm	M...
			5A	39	1 MWF	134 Comm	A...
			6A	43	2 MWF	134 Comm	B...
			7A	20	3 MW	134 Comm	D...
Ec 202	3	40	5A	40	1 MWF	137 Comm	S...
Ec 203	3	28	6A	28	2 MWF	137 Comm	S...
.							
.							
.							
Ec 501	3	3	9A	2	**	**	A...
			9B	1	**	**	M...
.							
.							
.							
Ec 507	3	32	0A	6	2-5:30 M	241 Comm	K...
			9A	11	3-5:30 W	241 Comm	S...
			9B	5	2:30-5 Tu	103 Comm	K...
			7A	10	3-5:30 Th	241 Comm	M...
.							
.							
.							

UNIVERSITY OF OREGON, EUGENE

Staff Member's Report of Service to the University

Fall Term, 19 61

Name John L. Doe Rank or Title Professor Dept. or School (and budgeted FTE) Economics 1.00

Office No. 103 Commonwealth Office Hours 10-12 MWF and by appointment

Original Signed by Staff Member ... Copies Signed by Dept. Head or Dean ...

This report covers basic data on the type and extent of services during a typical week, averaging out occasional services on a proportional basis (enrollments taken as of the end of the second week.) It is to be filled out by each person listed on an instructional division salary budget. It must be signed as indicated above, and filed (two copies) in the President's Office not later than the fifth week of each term.

TEACHING. Courses listed in the current Time Schedule to which time is given.

Dept. and No.	Sec.	No. Students Enrolled	No. Clock Hours in Typical Week Involved in This Course for: Lecture	Lab	Group Discussion	Individual Conference	Preparation and Papers	Total Hours in Typical Week
Ec 201	1A	56	3			3	3	9
Ec 325	2A	58	3			3	3	9
Ec 470	2A	28	3			3	3	9
Ec 503	9A	2				2		2
Ec 505	9A	8			3	1		4
Ec 507	9A	5			3		3	6

RESEARCH. Briefly describe any research project (subject and material, support, time involved).

American Labor Legislation — 5

ADVISING. Official assignment as faculty adviser for 18 undergraduates 9 graduates. — 2

OTHER SERVICES. Briefly describe (identify committees, organizations addressed, etc.); estimate the average weekly time involved. (Mention in parentheses if also on another OSSHE budget.)

Chairman, ... Committee — 5

Member, ... Committee — 1

TOTAL HOURS OF SERVICE TO THE UNIVERSITY IN A TYPICAL WEEK — 52

A 185

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

A CONCEPT OF THE MEASUREMENT OF FACULTY LOAD

CHARLES E. HOWELL
Northern Illinois University

FACULTY LOAD has long been limited to the narrow concept that the number of credit hours taught is an adequate measure of any individual staff member's entire work load. However important this aspect of the duties of a teacher may be, it is certainly not the whole job which he is called upon to do, and the mere number of hours of credit taught is certainly not an adequate measure even of that part of his job.

Any acceptable measure of total load must take into account not only all of those factors which the institution recognizes as part of the instructor's job, in the sense that it expects him to perform them even though they may not actually be included as a specific part of the contractual relationship, but it must also take into account those tasks which an individual feels are a burden upon him when he comes to perform them. Obviously then, there are far more factors in the total load picture than are reflected by any simple statement as to the number of hours taught.

Aspects of a Measurement System

Any method which seeks to broaden this measurement of total load must have at least two parts: 1) identification of the various aspects of load as recognized by both the institution and the faculty member and 2) weighting of the various aspects for comparability between faculty members. The measurement of these factors must be as objective as possible. The various factors must lend themselves to combination for comparison of one faculty member with another and to combination for comparison of departments. Combination of the various aspects must also be such that the different aspects of load can be identified within the entire institution. Finally, the method must be simple enough so that it can be easily understood and applied.

Faculty Load Measurement at Northern Illinois University

Identification and Weighting

This report illustrates how such a method has been worked out and applied at Northern Illinois University. The identification of the various aspects of load and the weightings assigned are shown in Chart I. Code numbers are given in this chart for easy reference to the tables which follow. The identification was worked out in conference with the academic vice president and the various deans and later was modified by suggestions of department heads. Certain weighting relationships already existed such as the relationship of undergraduate to graduate work, on-campus work to extension, and laboratory work to lecture courses. These relationships were carried over into the new weighting and extended.

Data were collected from each faculty member by use of the form of Chart II. The forms were collected through the department heads, who inspected and initialed them. The questionnaire of Chart II is filled in with illustrative data.

Computation of Individual Scores

Upon receipt of the questionnaires all data were checked insofar as possible with the class load schedule, the university committee list, and other objective university records. The method of computation of the individual score may be seen from the following illustration based on the data of Chart II.

Staff member XYZ has seven hours of undergraduate courses, W217, W316, W472. At 10 points for each hour he receives a credit of 70 points. He has the graduate course, W561, for two hours, for which he receives 30 points at 15 points for each hour. He has the course, W472, for three hours taught in extension. He receives 15 extra points for this. He teaches W217 as a laboratory course meeting two extra laboratory hours weekly, for which he receives 10 points.

He drives 105 miles to the extension course and receives two points for each additional 10 miles or fraction thereof over 80 for a total of six points. In W316 he has 43 students which is an excess of 13 above the 30, for which he receives one-fourth point each or a total of three. He has somewhere between one and 10 undergraduate advisees, for which he receives five points. He has the same number of graduate advisees, for which he receives 10 points.

He has one person enrolled in W497 for independent study for which he receives three points. He is

CHART I

POINT SCHEDULE FOR STAFF LOAD*

Code	Points	Aspect of Load
A	10	Each hour of undergraduate work taught
B	15	Each hour of graduate work taught
C	5	Additional points for each hour taught in Extension
D	5	Additional points for each laboratory hour taught (7 points for graduate laboratory)
E	2	Additional points for each 10 miles or fraction thereof for mileage in Extension over 80, but not more than 10 points
F	$\frac{1}{4}$	For each individual enrollment over 30 in each class, not cumulative from class to class - rounded to nearest whole for total
G	5	For each 10 or fraction thereof undergraduate advisees
H	10	For each 10 or fraction thereof graduate advisees
I	3	For each individual enrolled for independent study (private lessons in music: 3 points per individual for $\frac{1}{2}$ hour course, 5 points for 1 hour course)
J	5	Each graduate paper currently directed
K	10	Each hour of released time for whatever purpose
L	2-8	For membership on each University committee
M	10	For each University committee chairmanship held (standing committees only)
N	2-10	For holding office in a state or national organization
O	10	For each additional preparation beyond 3 (Except for activity courses where 3 points are assigned)

*There is no duplication if time is released for committee work or because of state or national offices held. Points are assigned for the released time and not for the committee work.

CHART II

INSTRUCTIONAL LOAD FIRST SEMESTER 1960-1961

NAME XYZ DEPARTMENT W

TEACHING DUTIES

Course and Section Number	Credit Hours	Hours Met Weekly	Enroll-ment	Miles Traveled Weekly if Extension Course
W561 - 1	2	2	21	
W217 - 2	2	4	17	
W316 - 1	2	2	43	
W472	3	3	29	105
W497 (IND. STUDY)	3	--	1	

ADVISEES
(check one only in each column)

Number	Undergraduate	Graduate
0		
1-10	X	X
11-20		
21-30		
31-40		
41-50		

NUMBER OF GRADUATE PAPERS which you are currently actively directing: 1 (Those of candidates scheduled to get the degree by August, 1961. If none, write "none.")

RELEASED TIME

Specific duties for which you are given release from teaching (include department headships, athletic coaching, debate, dramatics, consultant services, regularly recurring conferences, etc.)

Duty	Hours Released
RESEARCH	2

COMMITTEES

University committees to which you are assigned. Underline if you are chairman. Do not include departmental committees.

GRADUATE COUNCIL

INTERNAL BUDGET COMMITTEE

FEE ASSESSMENT COMMITTEE

OFFICES HELD

National or state organizations of which you are an officer. Indicate organization and office and approximate number of hours required per week through the current semester.

VICE-PRESIDENT, NATIONAL W...TEACHERS ASSOCIATION 2 HRS./WK.

STATE CHAIRMAN, LEAGUE OF INDEPENDENT VOTERS

currently directing one graduate paper for a candidate who is expected to get his degree prior to September, 1961, for which he receives five points. He is released officially from two hours of teaching for research, for which he receives 20 points.

He is a member of the Graduate Council, which is a major standing committee, meeting monthly, for which he receives five points. The Fee Assessment Committee, of which he is a member, serves only for a few hours during the registration period at a time when no classes are scheduled, and hence he receives no credit for this. He is the chairman of the Internal Budget Committee, for which he gets 10 points. He also receives no credit as state chairman of the League of Independent Voters since this is a civic and not a staff responsibility. He is, however, the vice president of a national teachers association which gives him three points if the work is considered to be spread over the academic year. He has four preparations exclusive of his independent study, for which he has received specific credit; hence he gets 10 points for the additional preparation beyond three. All of this gives him a total of 205 points.

Individual scores were collected by the department in which the individual served. If he taught in more than one department, the score was apportioned between the departments involved although his major score was assigned to the department to which he was contractually assigned. Table I gives a complete departmental listing, together with the footnotes which explain the point assignments. These complete footnotes were furnished to all department heads and academic deans and were available for inspection by the faculty members within a given department. Each department head was given the open lists of his own department but only coded names for members of other departments.

Load Summaries

A sample of the load summaries by selected departments in terms of point scores is shown in Table II. The same information by percentages by selected departments is shown in Table III and in graph form for the university as a whole in Chart III. Mean loads of full-time staff are shown in Table IV.

The "Normal" Load

In attempting to arrive at some kind of a norm, the mean points of all full-time instructional staff were calculated. Part-time staff were not included in the establishment of the individual or departmental mean load as a total. However, the work of part-time staff members was included in computing the total load of the department in its various aspects as well as the total load of the university as a whole.

Since the mean of the individual point scores is admittedly subject to variation, the normal load was defined as being a range of point scores extending one probable error on either side of the mean of the total distribution. This central range statistically encompasses 50 percent of the scores. Table V shows by department the number of individuals on a full-time basis who fell within this 50 percent (25% on either side of the mean) as well as those who fell from one to two probable errors on either side of the mean, roughly 16.13 percent in each case, and those who fell more than two probable errors on either side of the mean, 8.87 percent in each case.

Reference to this table and to the detailed departmental breakdowns will enable the identification of individuals and/or departments with varying degrees of over-and under-load.

Results of the Study

In extreme cases there may be aspects of load factors which are not uncovered by consultation either with the staff member concerned or with the department head. Therefore, the extremes particularly should be viewed with caution. Also, departmental averages, particularly in the case of newly established or very small departments, must be considered in the light of these aspects of load factors. However, it is believed that on the whole this procedure has objectively identified the factors entering into the total load of the instructional staff of the University. The study also throws light on the distribution of recognized aspects of load throughout the university which could be, although they have not been at Northern Illinois University, translated into cost estimates.

TABLE I

POINT ASSIGNMENTS FOR FACULTY OF "A" DEPARTMENT

Code Number	Aspect of Load															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Total
0065	140	--	--	25	--	8	15	--	--	--	--	--	--	--	10	198
0243	60	90	--	36	--	--	--	10	--	--	--	5	--	--	10	211
0310	60	--	--	15	--	--	15	--	3	--	60	4	--	5	--	162
0360	60	--	--	30	--	--	--	10	--	15	60	5	--	3	--	183
0370	35	45	--	--	--	44	--	20	--	10	30	--	--	--	--	184
1070	120	--	--	20	--	9	10	--	--	--	--	--	--	3	--	162
1297	110	--	--	30	--	6	--	--	--	--	--	--	--	--	--	146
1302	100	--	20	30	10	--	--	--	--	--	--	--	--	--	--	160
2230	90	--	--	45	--	--	10	--	--	--	20	2	--	--	--	167
2261	105	--	--	40	--	44	--	--	--	--	--	--	--	--	--	189
2373	60	--	--	30	--	--	--	--	--	--	30	--	--	--	--	120
2540	90	--	--	30	--	--	--	--	3	--	--	3	--	--	--	126
TOTAL POINTS	1030	135	20	331	10	111	50	40	6	25	200	19	--	11	20	2008
AVERAGE POINTS																166.7
% of Total	51.3	6.8	1.0	16.5	.5	5.5	2.5	2.0	.3	1.2	10.0	.9	--	.5	1.0	100.0

1. 0243 receives 5 points as a member of the Graduate Council.

2. 0310 has 3 hours of released time as director of the Art Gallery and 3 hours for work in Regional Services. Receives 2 points as a member of the Homecoming Committee, 2 points as a member of the Instructional Materials Advisory Committee, and 5 points as the editor of I.A.E.A. Directions.

3. 0360 has 6 hours of released time as chairman of the General Education Advisement Program. Receives 3 points as a member of the Teacher Education Advisory Council, 2 points as a member of the Instructional Materials Advisory Committee, and 3 points as chairman of the editorial board of the Illinois Art Education Association.

4. 0370 and 2261 teach section 2 of Art 214 together. The points for the course are divided equally between them.

5. 0370 has 3 hours of released time as head of the department.

6. 1070 receives a total of 3 points as scholarship chairman of Delta Kappa Gamma and as secretary of the local AAUP chapter.

7. 2230 has 2 hours of released time for making slides for lectures on the history of art. Receives 2 points as a member of the Artists and Lecture Board.

8. 2373 has 3 hours of released time during which he is assigned to the Dean of the College of Fine and Applied Arts.

9. 2540 receives 3 points as a member of the General Education Advisement Committee.

TABLE II

DISTRIBUTION OF POINT SCORES BY ASPECT OF LOAD AND BY SELECTED DEPARTMENTS

Department	Aspect of Load															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Total
Art	1030	135	20	331	10	111	50	40	6	25	200	19	--	11	20	2008
Biological Sciences	1620	105	--	219	10	267	160	110	18	40	170	21	--	10	50	2800
Bus. Ad. & Bus. Educ.	3060	195	20	75	10	131	485	140	69	175	220	40	--	21	40	4681
Chemistry	860	30	--	272	--	27	113	75	3	10	80	25	15	9	--	1519
Earth Science	930	90	--	90	--	49	80	90	--	50	210	20	--	8	--	1617
Education	3170	1170	120	475	26	58	415	810	--	500	410	75	10	30	20	7289
Etc.																
TOTAL	29205	3315	296	3110	114	1792	2795	2110	908	1205	5454	464	115	194	599	51676

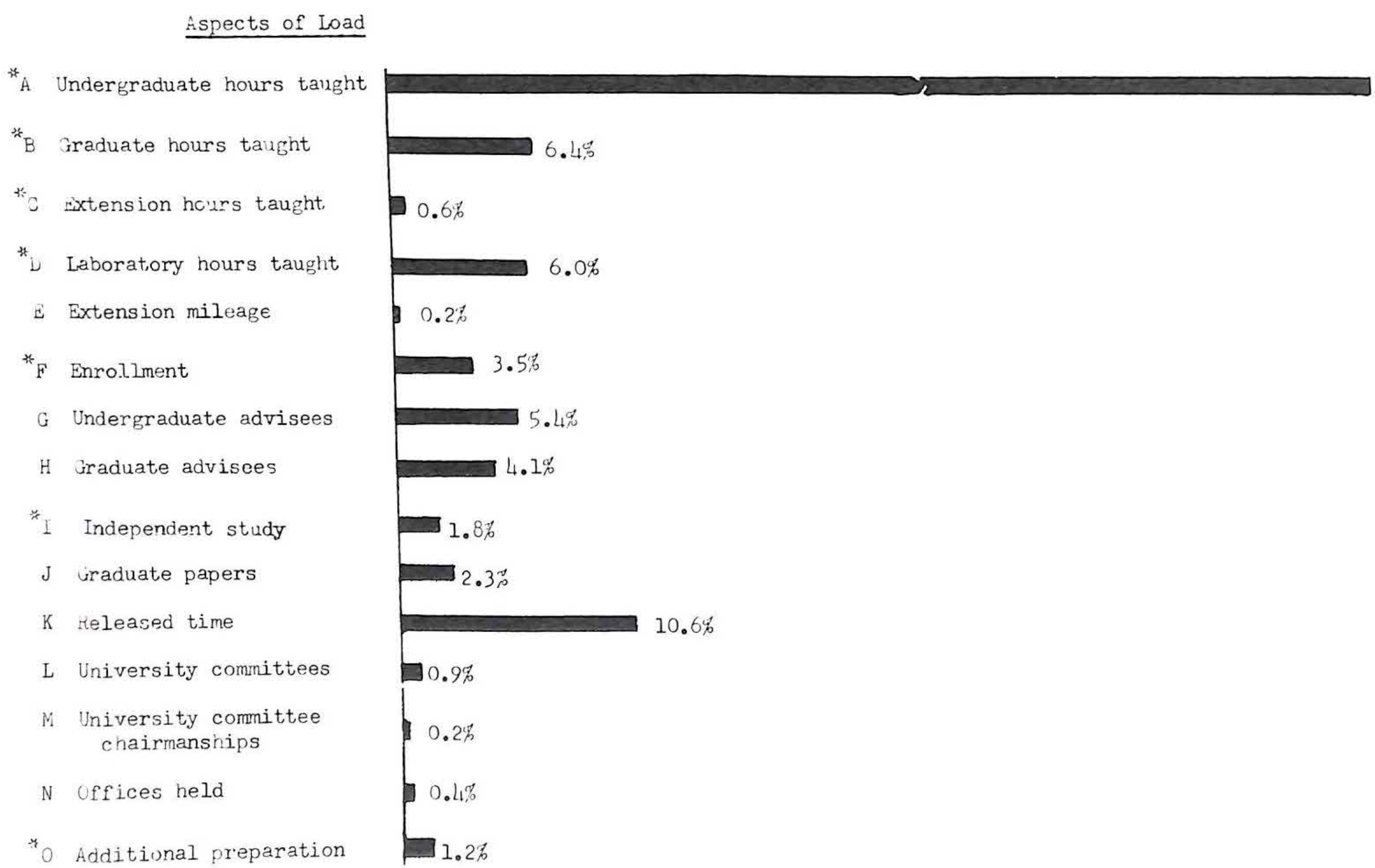
TABLE III

DISTRIBUTION OF ASPECTS OF LOAD BY PERCENTAGE OF LOAD
BY SELECTED DEPARTMENTS AND FOR THE UNIVERSITY AS A WHOLE

Department	Aspect of Load														
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Art	51.3	6.7	1.0	16.5	.5	5.5	2.5	2.0	.3	1.3	10.0	.9	--	.5	1.0
Biological Sciences	57.8	3.8	--	7.8	.4	9.5	5.7	3.9	.6	1.4	6.1	.8	--	.4	1.8
Bus. Adm. & Bus. Educ.	65.4	4.2	.4	1.6	.2	2.8	10.4	3.0	1.5	3.7	4.7	.8	--	.5	.8
Chemistry	56.6	2.0	--	17.9	--	1.8	7.4	4.9	.2	.7	5.3	1.6	1.0	.6	--
Earth Science	57.5	5.6	--	5.6	--	3.0	4.9	5.6	--	3.1	13.0	1.2	--	.5	--
Education	43.5	16.1	1.6	6.5	.4	.8	5.7	11.1	--	6.9	5.6	1.0	.1	.4	.3
Etc.															
All Departments Combined	56.5	6.4	.6	6.0	.2	3.5	5.4	4.1	1.8	2.3	10.5	.9	.2	.4	1.2

CHART III

DISTRIBUTION OF ASPECTS OF LOAD BY PERCENTAGE OF LOAD FOR THE UNIVERSITY AS A WHOLE



*Directly related to class instruction.

TABLE IV

MEAN LOAD IN POINTS OF FULL-TIME STAFF
BY SELECTED DEPARTMENTS AND FOR ALL DEPARTMENTS COMBINED

Department	Number of Full-Time Staff	Total Point-Load of Full-Time Staff	Mean Point-Load of Full-Time Staff
Art	12	2008	166.7
Biological Sciences	14	2800	200.0
Bus. Adm. & Bus. Educ.	23	4421	192.2
Chemistry	8	1390	173.8
Earth Science	9	1617	179.7
Education	38	7153	188.2
Etc.			
TOTAL	283	49722	175.7

TABLE V

DISTRIBUTION OF FULL-TIME STAFF BY SELECTED DEPARTMENTS
IN TERMS OF LOAD AS RELATED TO PROBABLE ERROR DISTANCE
FROM THE MEAN LOAD OF ALL FULL-TIME STAFF

Department	Within 1 P.E. of the Mean (middle 50%)	Below the Arithmetic Mean		Above the Arithmetic Mean	
		1 P.E. - 2 P.E. (16.13%)	2 P.E.+ (8.87%)	1 P.E. - 2 P.E. (16.13%)	2 P.E.+ (8.87%)
	Score 152-199	Score 128-151	Score below 128	Score 200-223	Score above 223
Art	8	1	2	1	--
Biological Sciences	4	--	1	7	2
Bus. Adm. & Bus. Educ.	13	3	1	3	3
Chemistry	6	1	--	1	--
Earth Science	7	1	--	1	--
Education	18	5	2	3	10
Etc.					
TOTAL	158	44	23	33	25

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

EDUCATIONAL FORECASTING AT THE UNIVERSITY OF AKRON*

D. J. GUZZETTA
The University of Akron

"If a man does not know to what port he is sailing, no wind is favorable."
Socrates

THE FORECASTING of educational needs for an institution of higher education is a complex and variable matter. Yet, the college or university that fails to maintain such a practice on a continuing and organized basis will find itself in difficulty as larger enrollments reach the doorsteps of our educational institutions beyond the high school.

The University of Akron, recognizing the need for such planning, both from a short and long range point of view, launched into an extensive analytical study of this type in 1957. Since then, detailed projecting and planning have gone on side by side with the end result being a well-informed faculty, staff, administration, and Board of Directors.

Effective planning for increased activity cannot leave any stones unturned. Consequently, The University of Akron examined several areas of institutional concern. It was agreed, however, that the point of departure for the study should be the potential and expected enrollment through 1970. On the basis of detailed projections arrived at through a purposive consideration of existing procedures and practices, requirements in curricula, physical facilities, finances, size and type of faculty, instructional techniques, student personnel services, library services and administration were studied. This was accomplished by local staff and selected national experts in particular fields. The end results taken as a whole have served as a "blueprint" for present and future institutional planning.

To illustrate the nature of the research involved as well as some of the techniques utilized, particular attention will be devoted to how enrollment, faculty and physical facilities needs were forecasted.

Enrollment

Forecasting enrollments at The University of Akron began with an intensive study of enrollments over the past 20 years. A number of pertinent facts were revealed as available data were analyzed. For example, enrollment breakdowns showed that almost consistently approximately 90 percent of the total day and evening students commute to and from the University. Also, since a large number of the students work at outside jobs while in attendance, average student academic loads are slightly below normal. The end result is that many students take more than the usual eight semesters to complete their educational objective. It was also noted that while in general about 80 percent of each entering freshman class is made up of graduates from Akron area high schools, about 15 percent of the total freshman student body finished high school prior to the previous June.

With these data as background, the school systems in the Akron area were solicited for information regarding their grade-by-grade enrollments and their predicted individual survival percentages class by class. Public, private, and parochial schools were involved. Also, non-Akron schools were studied in terms of past enrollment practices. On the basis of the statistics collected plus prior patterns of detailed enrollment figures at the University, future entering day freshman classes were predicted from 1958 to 1970. These projections were made on the assumption that a number of existing conditions, practices, and patterns of operation would continue to be applicable in the overall picture.

---

* For detailed specifics regarding techniques used and depth studies conducted, please contact the writer. He will be pleased to expand in any of the areas described above or alluded to in this article.

Being cognizant of the many variables that enter into an enrollment forecast of this type, four freshman projections were made, each based upon specific assumptions. In this way, the more pertinent considerations were appropriately grouped into logical patterns. The end result was that forecasts were made on the basis of clear and explicit basic assumptions. Without such a basis, forecasting becomes mere guesswork and speculation.

Certain assumptions seemed applicable to all four projections. They were arrived at after careful study of the local situation and other student-producing areas outside of Akron. In brief, these assumptions dealt with the reliability of the data provided the University by the various high schools, the continuing attraction to the University of regular quotas of students, the increased interest on the part of high school graduates to attend institutions of higher learning in general and The University of Akron in particular, the supposition that no other institution of higher education will be established in Akron, and the probability that the same percentage (15%) of returnees, transfers, and entering freshmen from classes other than the immediately preceding June high school graduating class would be maintained.

The first (number 1) of the four projections was deliberately developed as minimal. Basically, it was structured assuming no percentage change in higher education interests, no change in University policies, fees or curricula, but with a slight increase in enrollment due to the presence of more graduates.

Projection number 2 included the same assumptions referred to under number 1, with the exception that the Akron youth population interested in attending The University of Akron will follow national predictions with a one percentage point increase annually through 1970.

The next projection (number 3) took into consideration the possibility of extending the University influence beyond the city limits of Akron. For example, even though The University of Akron was originally established as a municipally oriented institution and its primary source of income comes from the city itself, its financial base could be enlarged to include outlying taxing districts. In effect, this could mean a slight reduction in the present differentiation in fees that exists between residents and non-residents. This, then, would attract a larger percentage of non-residents and thereby add to the University student body.

Under projection number 4, this financial base is assumed to be further expanded to the point where the University would become completely county wide and in this way eliminate, to a large degree, the present fee differential. Also, new curricula would be added to serve a larger population base. Under this plan, the increase in student interest in higher education at The University of Akron would rise by approximately one percent per year throughout the county as a whole. Under number 3, however, the increase would amount to about one-half of one percent per year. Table I identifies the freshman day totals developed through these four projections.

To arrive at total day enrollments, four projections also were developed using the same assumptions described for freshman totals. On the basis of the official record of total enrollments over the past 20 years, the following formula was applied in each projection.

Total enrollment = 85% of the three previous Fall freshman classes plus the projected freshman enrollment for the particular year.

This was validated by applying the formula to every academic year back to 1937 and in no instance was the variation greater than 25. Table II lists total projected day enrollments based on this procedure. Table III shows the forecasted evening enrollment figures. The latter were arrived at by using "A" and "B" projections. The "A" forecasts represent approximately 90 percent of expected day enrollments while the "B" column assumes a 95 percent total. In the judgment of the University, Column "A" will probably be more nearly accurate.

Faculty

The primary function of a university is to create an atmosphere favorable to and provide the means for learning at an academic level beyond the high school. To do this effectively, an institution of higher learning must be endowed with a top level, able faculty.

It is no secret that obtaining the academic services of well-qualified faculty is going to be a major area of consideration for the future. Without such personnel, a university has little reason for being. Therefore, faculty needs must be determined early in order to be prepared to carry on the various educational services. This important area of faculty needs, then, loomed as a major point of examination in the forecasting of required short and long range personnel resources.

The University of Akron arrived at a determination of faculty needs by projecting the student credit hour demands. Since at this institution student costs are arrived at on the basis of the number of credit hours a person takes, an average student credit hour load was used to project student credit hour demands over the next several years. The procedure was a particularly significant one, inasmuch as the average semester student credit hour load is approximately 14 hours, and so faculty,

TABLE I

FRESHMAN DAY FALL ENROLLMENT PROJECTIONS 1958 - 1970

Year	Projection Type No. 1	No. 2	No. 3	No. 4
1958	630	690	650	650
1959	675	720	735	905
1960	910	970	1005	1125
1961	810	1025	1065	1330
1962	810	1050	1120	1420
1963	820	1090	1175	1520
1964	940	1300	1390	1835
1965	1040	1485	1605	2165
1966	1050	1500	1700	2200
1967	1025	1550	1775	2300
1968	1000	1500	1800	2350
1969	1100	1650	2000	2600
1970	1250	1800	2200	2900

TABLE II

TOTAL DAY FALL ENROLLMENT PROJECTIONS 1958 - 1970
(Rounded to nearest five)

Year	Projection Type No. 1	No. 2	No. 3	No. 4
1958	2440	2500	2460	2460
1959	2400	2500	2465	2645
1960	2850	2900	2950	3075
1961	3150	3250	3300	3550
1962	3300	3600	3700	4240
1963	3400	4000	4100	4950
1964	3425	4225	4500	5700
1965	3625	4650	5020	7190
1966	3800	4800	5200	7300
1967	4000	4900	5350	7550
1968	4050	5000	5500	7800
1969	4100	5100	5600	8000
1970	4200	5400	5900	8500

TABLE III

TOTAL EVENING CREDIT FALL ENROLLMENT PROJECTIONS 1958 - 1970
(Rounded to nearest five)

Year	No. 1		No. 2		No. 3		No. 4	
	A	B	A	B	A	B	A	B
1958	2200	2320	2250	2375	2210	2350	2210	2350
1959	2150	2275	2250	2375	2225	2350	3400	2500
1960	2200	2700	2300	2750	2350	2800	2700	2900
1961	2850	3000	2950	3100	3125	3225	3200	3375
1962	3000	3150	3250	3450	3350	3525	3800	4000
1963	3050	3225	3550	3725	3700	3900	4450	4700
1964	3075	3250	3850	4025	4050	4275	5150	5425
1965	3275	3450	4200	4425	4500	4775	6475	6825
1966	3425	3500	4350	4550	4700	4850	6570	6950
1967	3600	3700	4410	4650	4815	5000	6795	7100
1968	3645	3750	4500	4750	4950	5100	7020	7300
1969	3690	3800	4600	4850	5040	5200	7200	7500
1970	3780	3900	4900	5000	5310	5400	7650	7900

TABLE IV

PROJECTED DEMAND FOR FULL-TIME FACULTY EQUIVALENTS
1958 - 1970 FALL SEMESTERS

Year	Student Credit Hour Demand			Demand for Full-Time Faculty Equivalents
	Day	Evening	Total	
1958	33, 750	10, 575	44, 325	141
1959	33, 669	10, 552	44, 221	140
1960	39, 204	10, 843	50, 047	159
1961	43, 808	13, 724	57, 532	183
1962	48, 749	15, 275	64, 024	203
1963	53. 082	16, 633	69, 715	221
1964	57, 227	18, 025	75, 252	239
1965	62, 883	19, 702	82, 585	262
1966	64, 800	20, 304	85, 104	270
1967	66, 150	20, 727	86, 877	276
1968	67, 500	21, 150	88, 650	281
1969	68, 850	21, 620	90, 470	287
1970	72, 900	23, 030	95, 930	304

income and facilities are planned for on a less than normal load. In this way, a more exacting working base could be used in forecasting total needs.

For the purpose of the faculty analysis, it seemed meaningful to go a step further and to deal with forecasts of full-time faculty equivalents or FTFE. An FTFE was defined as a teaching load of 12 credits — a "normal load." Also, for purposes of faculty equivalent projections, both day and evening student credit hour demands were considered. This was arrived at as follows.

First, the years from 1952 to 1957 were selected as the base period from which to project. The actual student credit hours produced during this period were determined, and then the fall semester totals were divided by 12 to arrive at an index of the mean number of student credit hours per full-time faculty equivalent. The mean for the five-year period was 315 per semester, and this therefore was regarded as the measure of productivity of the faculty during these years. It was agreed that at least the same load level would prevail over the next several years.

The projected student credit hour demands arrived at on the basis of enrollment forecasts and average subject matter loads were then divided by 315 to arrive at full-time faculty equivalents required in the future for each fall semester. The results are shown in Table IV but only on the basis of student projection number 2. It should be noted that no attempt was made at this time to determine how many part-time faculty should be planned for. It seemed best to retain that flexibility for planning purposes each year. The availability of qualified full-time faculty members would determine the extent of part-time faculty participation in the total day and evening program.

Physical Facilities

A growing student body and a growing faculty require appropriate physical facilities that will aid both groups in achieving their educational objectives. Specifically, this means adequate classrooms, offices, library facilities, and dormitories. Accordingly, the point of departure for this phase of the University's forecasting task was the accumulation of factual information on the present utilization of existing facilities. This study was accomplished in accordance with a widely used manual on space utilization by Russell and Doi. *

As a result of this facilities study, the following information was assembled to provide a basis for detailed planning and analysis:

(1) Summary of square feet, student stations, number of rooms and average square feet of floor space per student station in the instructional facilities.

(2) Summary of rooms and student stations available during the week according to morning, afternoon, and evening use and percentage of use as well as by functions of rooms.

(3) Summary of the use of laboratories.

(4) Distribution of the number of day class period meetings per week by enrollment of classes in relationship to the capacity of the rooms in which current classes were held.

(5) Summary of percentage of station use when classrooms were actually in use during the academic year under study.

Volumes of statistical data resulted, and the findings were analyzed in terms of next steps to take as the availability of space becomes crucial. Such conclusions as the following provided the norm from which to build, inasmuch as operational efficiency and effectiveness was at an adequate level at the time of the study and for the current enrollment.

(1) The gross square feet of space on the campus averaged 200 square feet per full-time equivalent day student and 160 square feet per full-time equivalent day and evening student.

(2) The academic program was housed in twelve different buildings providing 3,316 student stations in general classrooms, 1,411 student stations in instructional laboratories, and 243 student stations in other instructional rooms for a total of 4,970 student stations. This amounted to an average of 19.2 square feet of assignable space per student station.

(3) Ninety percent of the classes that met in general classrooms were in facilities which had more student stations than class enrollment. In instructional laboratories, 73 percent of the classes met in facilities having more student stations than the class enrollment.

(4) When in use, general classrooms were 48 percent filled to capacity and instructional laboratories were 61 percent filled to capacity.

Certainly it is obvious that 100 percent space utilization for all types of facilities during a full week's university operation would be most difficult, if not impossible, to achieve. Scheduling complications for both students and faculty would provide the most serious obstacle to full utilization. Therefore, the existing pattern, with a wider range of class distributions, was projected into the future to determine what new facilities would be required

* John Dale Russell and James I. Doi. Manual for Studies of Space Utilization in Colleges and Universities, a service of American Association of Collegiate Registrars and Admissions Officers (Washington, D.C.: Publications Office, American Council on Education, 1957).

to assure a continued effective operation. The application of sound, reasonable, and practical principles in this building program followed.

As an illustration, it was discovered that only 415 actual student seating stations were available in the library. This represented less than 10 percent of the total student body, day and evening. According to experts, 20 to 50 percent of the total student enrollment is considered to be the ideal seating capacity in a well-planned library. Taking a minimal percentage of 20 percent of the total student body for planning purposes, this figure was projected into future needs including the introduction of more carrel space for study purposes. Also because of the substantial student increases expected, the expanded graduate program under consideration, and the existing cramped stack space available, there appeared to be no question but that the "intellectual center" of the campus also needed expanding for additional book stacks.

A compilation of all available data assisted the University in pinpointing its facilities needs on a priority basis. The conclusion reached was that with the exception of the library and a dormitory (a new development at the University), the increased enrollments anticipated between 1958 and 1961 could be accommodated provided tight and efficient use of space would follow. Building needs were then established in priority order as follows: (1) a library addition (1959); (2) a dormitory (1960); (3) a general classroom building (1962); (4) two more dormitories (1962); (5) a second general classroom building (1964); (6) an administration building (1966); (7) a special purpose instructional-laboratory type building (1968).

Conclusion

Experience has shown that institutional studies, such as the one reported here, conducted principally by local University staff are of great value to the institution. Sometimes they are more helpful than those accomplished by outside agencies. Among the advantages directed towards this internal approach are that campus personnel can establish their own objectives, determine the problems to be studied, and apply predetermined procedures and techniques to the actual operation of the study. The tasks or issues at hand are more important to the local staff than to outsiders; hence they are more intent upon achieving discernible results.

The opportunity to utilize local staff to assist in the accumulation and interpretation of data and the freedom to call in special consultants on a need basis would seem to result in more effective decisions. This is so since they are based on a combination of judgments made from the practical experiences of local staff and the varied experiences in higher education of the outside experts.

In a sense, every institution of higher education is unique, regardless of its many commonalities with similar type organizations. The communities in which educational institutions are located are different. Their physical setting and environment, although comparable, are unlike other seemingly identical institutions. The human and financial resources available in each situation vary. Each locale has its own idiosyncracies that need to be taken into account. Therefore, generalizations, in terms of acceptable procedures and solutions to problems, cannot be applied without becoming influenced by historical "accidents" responsible for much of the "status quo." At the same time, the experiences of the authorities in the field provide the basis for recommendations aimed at helping an institution of higher education maintain its individuality while holding on to its academic respectability.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

FACULTY SATISFACTIONS AND DISSATISFACTIONS

JOHN DALE RUSSELL
New York University

THE IDEA for a study of faculty satisfactions and dissatisfactions had been developed over a period of years at New York University by the Committee of the Elected (faculty) Members of the Senate of the University. After the establishment of the University's Office of Institutional Research in 1958, that Office agreed to take responsibility for carrying on the study. The Committee of the Senate continued to be in constant touch with the study throughout its development and gave valuable advice on the method and content to be used.

The study was conducted by means of a Questionnaire distributed to all full-time members of the New York University faculty in May, 1959. Respondents were asked not to sign the Questionnaire. Returns were received from only about 45 percent of those who were asked to report. A careful and critical analysis of the returns shows that they provide an excellent sample of the entire group of full-time faculty members, with respect to a number of personal characteristics on which comparisons between the sample and the entire faculty group could be made.

The Questionnaire

The Questionnaire* was long, consisting of 15 pages of items, in check-list form, to which reactions were asked. The items were divided into six major parts. Part I contained 11 items identifying certain characteristics of the respondents, such as the School or College of the University with which connected, academic rank, age, University salary, highest earned degree held, and years of service at New York University.

Part II consisted of a list of 88 items giving situations for which respondents were asked to indicate on a 9-point value scale their general feeling of satisfaction or dissatisfaction at New York University. The instructions were to check the midpoint on the scale, 5, if the condition mentioned in the item was one for which the respondent's satisfaction and dissatisfaction were approximately equal. The response-value of 5 is thus referred to as the neutral point on the scale. Scale values from 6 to 9 were to be checked to indicate varying degrees of satisfaction, with the highest response-value, 9, indicating the very greatest degree of satisfaction. Response-values of 4 and below similarly indicated varying degrees of dissatisfaction, with the lowest, 1, to be used for a serious condition affecting very adversely the respondent's effectiveness and morale.

Part III of the Questionnaire asked for reactions to 18 items of suggested innovations in University services, and for reactions to 10 suggested areas for uses to which any increased University income might be put. In the case of each item, the response-value scale from 1 to 9, again was used so that respondents might indicate the effect (negative, neutral, or positive) on their feeling of satisfaction at the University.

Part IV of the Questionnaire related to professional attitudes toward service at New York University, again using the 9-point scale for indicating positive, neutral, or negative attitudes toward the various areas of service that were listed. This section also asked for information concerning the scholarly activities of the faculty respondents and their connections with learned societies and professional organizations.

Part V of the Questionnaire related to the economic status of the faculty respondents. It included questions in such areas as housing, earnings of spouse, total income of self and spouse outside the University salary, and number of dependents.

Part VI related to attractions away from New York University. One set of questions asked for information about offers received from other institutions within the past three years. Inquiry was also made about the faculty member's attitude toward the encouragement of such offers. Respondents were asked to evaluate each of some 14 factors that might be present in an offer, with respect to the importance of the factor on the decision to leave or to remain at New York University. Request was made for the names of a few institutions to which the respondent would be attracted by an offer not involving a salary higher than that currently being received at New York University.

* A copy of the Questionnaire may be obtained by addressing the Office of Institutional Research, New York University, New York 3, New York. The Questionnaire was set up in a form to simplify processing by punched card tabulation.

The 9-point Scale

The purpose of setting up the items of the Questionnaire for response by checking, usually on the 9-point scale, was to facilitate tabulation by the IBM process and also to simplify the answering of the items by respondents. Opportunity was also offered on each item for the respondent to write in any comment he cared to make.

The 9-point checking scale worked out very well. The respondents tended to distribute their reactions among all the possible response-values. On the average, 26.6 percent of those responding to the 88 items of Part II checked the neutral response-value of 5, indicating that the item did not affect one way or another their feelings of satisfaction or dissatisfaction at the University. The other responses in Part II were distributed fairly evenly among the remaining eight response-values in the scale, though the general trend of responses differed considerably from item to item.

One striking fact brought out in the tabulations is the extent to which the faculty respondents differed among themselves in their reactions to each of the items in the Questionnaire. On every item in Part II and also in almost all items in other Parts of the Questionnaire, the highest and lowest extremes on the response-value scale were always checked by some respondents. This was not a result of any tendency for just one or two eccentric respondents always to check a response-value of 9 or of 1 on every item. Indeed, the respondents seemed to be very conscientious and discriminating in assigning response-values to each item in accordance with their personal evaluation of its effect on their feeling of satisfaction or dissatisfaction.

The tendency for both the highest and lowest response-values to appear in the reactions to each item was true, not only for the entire group of 580 respondents but also for practically every category into which respondents were divided, for purposes of analysis on each of four classification items--academic rank, age, salary, and School or College within the University. In any category having a substantial number of respondents, say 40 or more, almost invariably the very highest and the very lowest possible response-values on the scale would appear among the reactions.

The tendency of faculty members to differ among themselves, to "think otherwise" as someone has put it, has often been noted. It was interesting to see how consistent the responses to the Questionnaire were in this respect, with the extremes as well as all the in-between values in the satisfaction-dissatisfaction scale being found on practically every item in the Questionnaire. This finding has important implications for administrative actions for it becomes clear that any decision that is made or course of action that is carried out is almost certain to meet widely differing reactions among faculty members.

Responses to Situations Affecting Satisfactions

In Part II of the Questionnaire the 88 items were divided into six major sections. Section A, relating to physical plant facilities, consisted of 21 items. The average response-value for the entire group of respondents for the 21 items on physical plant facilities was 4.8, slightly on the unfavorable side of the scale. Classrooms, laboratories, research rooms, faculty offices, and elevator services were generally scored rather low. The item rating highest in this group was bookstore facilities.

Section B of Part II consisted of six items related to personnel services (clerical assistance, graduate assistants, janitors, library staff). The average response-value for this group of items was 5.2, slightly above the neutral point of the scale. In general, ratings were low for the <u>extent</u> of clerical and graduate assistance but above the neutral point for the <u>quality</u> of such assistants. The library staff received a rather high rating, indicating that this is in general a source of considerable satisfaction to faculty members.

Section C of Part II consisted of 17 items related to faculty personnel policies and practices. The general trend of reactions on these items was slightly unfavorable, with an average response-value of 4.8. All the items in this section relating to faculty salaries received low ratings on the average. The "fringe benefit" provisions, however, such as retirement, life insurance, health insurance, tuition remissions, etc., received generally favorable reactions. The item, "funds for expenses of travel to professional meetings," was given a very low rating.

Section D of Part II consisted of 30 items related to "other administrative practices and policies." The response-values on this section averaged out to 5.0, exactly the neutral point on the scale. Items related strictly to academic administration in general had an average response-value on the "satisfaction" side of the scale. Participation by faculty members in various kinds of administrative decisions was in almost every case rated low, indicating some substantial extent of dissatisfaction arising from the present level of faculty participation. A general trend was observed for the responses to be more favorable to administrative policies and practices at the departmental and School or College level than at the all-University level.

Section E of Part II consisted of six items relating to faculty colleagues. Here the picture was one of rather general satisfaction; the average response-value was 6.3, the highest for any of the major sections of Part II. It is evident that for the average faculty member, his colleagues on the faculty are an important source of satisfaction in his service at New York University.

Section F, the final one in Part II of the Questionnaire, consisted of eight items related to the characteristics of students. The average response-value for the group of items in this section was 5.2, slightly on the favorable side of the satisfaction-dissatisfaction scale. The average response-value for "the previous preparation of your students" was only 4.7, a little below the neutral point on the scale. Average response-values well above the neutral point were given such items as "the seriousness of purpose of your students," "the general promise of the students in this University," and "the extent to which students are achieving in accordance with their abilities."

When the responses to the 88 items in Part II of the Questionnaire were analyzed in accordance with the Schools and Colleges of the University with which the respondents were affiliated, rather wide differences appeared in the average response-values for the various units of the University. Some of these differences served as a validation of the authenticity of the Questionnaire responses. For example, one School or College, that was most unsatisfactorily housed at the time the study was made, yielded a very low average response-value on the section related to physical plant; the condition was corrected shortly by the construction of an excellent new building. Other differences in average response-values for some of the Schools and Colleges, according to competent observers, reflected rather accurately the considerable differences that were then existing in the level of faculty morale in the various units of the University.

When respondents were classified according to academic ranks, the spread of average response-values was much less than when the classification was made according to School or College. In general, the lowest average response-values tended to come from the associate professors and the highest from the (full) professors, with instructors and assistant professors occupying an in-between position.

When the responses were analyzed according to the ages of respondents, only relatively small differences appeared in the average response-values for the various age-groups. In general, the older respondents tended to report somewhat higher levels of satisfaction than the younger, but the differences on most items were not large.

Analysis of the responses according to the University salaries of the respondents resulted in a wider spread of average response-values than in either of the classifications on the basis of academic rank or age. This was largely because of the tendency of those in the highest salary-brackets to report relatively higher levels of satisfaction than the total group of respondents. One could readily reach the general conclusion, on the basis of the statistical data, that one way to keep at least half the faculty in a state of satisfaction would be to make the minimum salary $10,000. Respondents in the highest salary-bracket, $15,000 and over, seldom reported any dissatisfaction with any condition whatever among the 88 items listed for response in Part II of the Questionnaire. In this group with the highest salaries, the responses to most items were enthusiastically on the "satisfaction" side of the scale.

Reactions to Suggested Innovations

Part III of the Questionnaire asked for reactions to a prepared list of suggested innovations in University services for faculty members. In the list were a number of suggested referral services to specialists on the University staff. The average response-values to all these were on the favorable side of the scale, with the highest average response-values being given to referral services on physical and mental health problems, legal matters, financial problems, and income-tax matters. Other forms of fringe benefits also generally drew responses on the favorable side of the scale, though for most of them the reaction was not particularly enthusiastic. The items drawing the highest average response-values were "tuition grants for children of faculty going elsewhere to college," and "University-owned rental housing near the campus."

A second section of Part III asked, "Under existing conditions at this University how could any given amount of increased income best be used to improve faculty satisfactions and morale?" A list of 10 items was presented for checking on the usual 9-point scale. All but one of the items received responses well on the favorable side of the scale. The item attracting the highest average response-value was "direct improvement in faculty salaries," with an average response-value of 8.3, extremely high on a 9-point scale. By contrast, the item "additional fringe benefits" rated much lower, with an average response value of 6.5, though still strongly on the favorable side of the scale. The item with the second highest average response-value, 7.1, was "improvement of facilities for research." Two others rating high were "improvements and renovations of the existing academic plant," and "improvement in the holdings and services of the University library." The one item that drew an average response-value on the negative side of the scale was "employment of a substantial number of additional faculty members."

Professional Attitudes

Part IV of the Questionnaire was concerned with professional attitudes and activities of the respondents. One section asked for reactions concerning the relative importance of nine different kinds of responsibilities for New York University. The usual 9-point scale was provided for indicating

importance (values of 6 or above) or unimportance (values of 4 or below). High average response-values were reported for several of these categories of responsibilities. The very highest, with an average response-value of 8.3, was "the teaching of graduate students." Only slightly lower, average response-value of 8.1, was "the teaching of undergraduate students." A close third was "advancement of knowledge through research," with an average response-value of 8.0. Tied for the same response-value was "maintenance of conditions within the University that are comfortable and attractive for excellent scholars." Other items ranking lower, with their average-response values, are: "enhancement of the reputation of this University as a center for higher learning," 7.8; "maintenance of a scholarly atmosphere within this University," 7.6; "preservation of the cultural heritage," 6.8; "application of knowledge to life situations," 6.0; and finally, "the solution of problems of great national and international concern," 5.6.

Another section of Part IV presented a listing of groups to which a faculty member might have some feeling of obligation for effective service at the University. This was also to be answered on a 9-point scale, in terms of the importance or unimportance of each group in the faculty member's attitude toward his obligation for effective service at the University. The list of groups, with the average response-values for each, is as follows:

The students in the faculty member's classes...	8.5
One's colleagues on the faculty in his own department	7.5
Scholars throughout the world in the faculty member's special academic field........	7.0
One's colleagues on the faculty throughout the entire University..................	6.3
Intellectuals in general in the United States and throughout the world................	6.1
The administrative officers of the University	5.8
The Board of Trustees of the University..	5.1
The alumni of the University............	5.0

The analysis yields evidence that faculty members in general are oriented somewhat more strongly to their subject-matter disciplines than to the institution in which they serve.

Economic Status

Part V of the Questionnaire related to the economic status of faculty members. About 46 percent reported home ownership, and an equal number reported they live in a rented house or apartment. About one-third of the respondents reported some annual income of spouse from employment. The average annual income of self and spouse from sources other than the respondent's University salary was reported as $4,523, but 18 percent reported such outside income as zero. The average number of dependents was two; 20.8 percent reported no dependents. About 18 percent of the respondents reported one or more dependents attending college.

Attractions Away from the University

Part VI of the Questionnaire inquired about attractions that might take faculty members away from New York University. More than half the respondents (55.6%) reported they had had one or more offers of employment elsewhere within the three-year period prior to filling out the Questionnaire but had rejected such offers in order to remain at New York University. About 17 percent were considering an offer at the time the Questionnaire was returned or had already decided to accept an offer elsewhere. It is interesting to note that about one respondent in six was poised for flight. One-eighth of the respondents indicated that they were actively seeking opportunity for a position elsewhere, and about half of all respondents reported that they would welcome such an offer.

One section of Part VI presented a list of 14 factors to which the respondents were asked to react in terms of the importance of each factor in their evaluation of an offer that might come to attract them away from New York University. The usual 9-point scale was used for the evaluation, on which the scale-value of 5 represented a neutral reaction; the values above 5 represented positive attractions away from New York University, and those below 5 represented negative reactions to the factors. The list of factors, with the average response-value assigned each, is given below, in rank order of the average response-value.

The scholarly reputation of the institution making the offer..........................	7.5
Substantially larger salary than that now being received........................	7.4
Opportunities for research..............	6.9
The extent of the normal teaching load.....	6.4
The kind of library facilities available.....	6.3
The kind of classroom, laboratory, and office facilities........................	6.2
The kind of housing available for the family	6.1
Fringe benefits, other than salary, retirement, etc.	5.7
Educational opportunities for children in the family............................	5.6
Opportunity to take on greater responsibilities and to render a greater service to society...........................	5.3
Opportunity to be near relatives or other friends............................	4.0
Opportunity to live in a different part of the country........................	3.9
Opportunity to live in a different city or town............................	3.8

Opportunity to get out of university work and into some non-academic organization..... 2.6

Again it may be pointed out that "fringe benefits" rated considerably below "larger salary" in the rank order list of factors that would attract faculty members to some other institution.

The final item of Part VI asked respondents to list the names of a few institutions of higher education to which they would be attracted by the offer of a salary only equal to or a little larger than that now received from New York University. Only 44.7 percent of the respondents mentioned any other institution in response to this question. Presumably the other 55.3 percent would not be attracted away from New York University unless the offer involved a larger salary than at present received. Many interesting observations can be drawn from the analysis of the responses when the respondents are classified in various ways. For example, those whose principal service at New York University was in research mentioned some other institution more frequently than those whose principal service was teaching or administration. The list of names of institutions and the frequency with which each was mentioned are very interesting and somewhat surprising. This list may be considered a reflection of New York University's place in the "pecking order" among institutions of higher education in the U.S.A., from the point of view of the University's own faculty members. The list affords those in the University who are responsible for academic administration some idea of the sources from which competition for capable staff members will be most difficult to meet.

The Detailed Report of the Study

This article presents only a much condensed summary of the Study of Faculty Satisfactions and Dissatisfactions at New York University. A more detailed analysis, providing upwards of 1,500 tables of data, has been prepared in typewritten form. Copies are on file in the Office of Institutional Research, New York University. The detailed report was prepared only for the use of the administration and faculty members of the University. Access to the detailed report is controlled by the Committee of the Elected Members of the Senate of the University, by whom the study was originally planned, and for whom it was made. A copy of the blank Questionnaire form may be obtained by an interested person on application to the Office of Institutional Research, New York University, New York 3, New York.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

PLANNING EDUCATIONAL FACILITIES

FREDERICK E. SCHWEHR
Board of Regents of Wisconsin State Colleges

THE PROBLEM of planning educational facilities is not merely one of obtaining "X" number of dollars to house "Y" number of students. Neither is it purely one of efficient utilization of space. It is a problem of envisioning, designing, and perfecting an environment for the students and faculty which will stimulate and perpetuate the learning process.

One has merely to look at industry to gain an insight into this approach. Industry's objective is to place its product in every home in the country. Industrialists begin by team-gathering and analyzing all the related data affecting the marketing of their product. They know, for example, the type of environment which will best house the product and most attractively present it to the public. They plan with intelligence and imagination the quality of enclosures which will enhance their product, perpetuate its use, and stimulate the curiosity of the uninformed.

If one wishes to better the purposes and ideals of higher education, the curricula must be presented and housed in such a manner that institutional objectives will be attained in an environment that will attract and stimulate youth and faculties in their search for knowledge.

In planning educational facilities in Wisconsin, an attempt is being made to study and correlate as much relevant data as possible. Three kinds of studies which have aided in the planning of educational facilities are the survey of physical facilities, facilities quality study, and the analysis sketch. These will be presented in that order.

The Physical Facilities Survey

The State Coordinating Committee for Higher Education published the first space utilization study based on the fall semester 1956-57 in April, 1958. The second study, based on the fall semester 1958-59, was published in April, 1960; a study is underway based on the fall semester 1961-62.

The utilization of space was studied for the following reasons: (1) to determine the present utilization of existing facilities, (2) to determine how utilization might be improved, (3) to pinpoint institutional strengths and weaknesses in building utilization, (4) to provide background data for future plant expansion based on need, (5) to provide information relative to space allocation standards, (6) to use the findings as a guide in the development of an integrated building priority list for the University of Wisconsin and the State Colleges, and (7) to continually keep abreast of the rapidly changing educational scene.

Pitfalls

The physical facilities survey must not be considered as the final answer to all problems confronting institutions of higher learning as they attempt to meet tidal wave enrollments, even so far as building needs are concerned. Such a study must be considered as a guide which, when correlated with other information, will give an objective approach to the solution of space problems. Some other items which must be considered along with the physical facilities data are the types of programs offered as well as those planned for the future, institutional objectives, scheduling, availability of funds, land access, past and present enrollment trends, projected enrollments, the availability and utilization of college teaching resources, and the quality of the space in terms of its intended educational purpose.

Institutions of higher education are often criticized for inefficient use of existing facilities. Studies have been quoted as to the utilization rate of instructional areas in colleges and universities. In many instances, terms and utilization figures have appeared without definitions. Data presented in such a manner may be misleading and are often useless. Without knowledge of the meaning or derivation of these figures it is extremely difficult, if not impossible, to make institutional comparisons. In addition, little mention is made of the necessity for examining all other relevant data.

Hard and fast standards of utilization and space requirements which may be applied to all institutions across the nation do not exist. The physical facilities of one institution cannot be compared with another without examining the curricula and institutional aims. An example of two different types of institutions would be to compare Wisconsin State Colleges at Stout and Eau Claire. Eau Claire State College primarily offers a liberal arts curricula, whereas Stout State College specializes in teacher education in the specific fields of Home Economics, Industrial Education, and Vocational Education.

Clearly, these institutions cannot be compared

in terms of educational objectives. This, in itself, will affect the needs in terms of size and type of physical facilities required. In addition, a look at The University of Wisconsin indicates that, like any other university, it is composed of many different colleges each with specific aims and, in many instances, different requirements of physical facilities. Graduate programs require a large number of research rooms and seminar-type rooms. This is in direct contrast to the lower division of undergraduate students who may often be instructed in large general groups in a small number of rooms. It is evident that broad generalizations on utilization standards cannot be applied. The specific needs of the curricula offered at each institution must be considered in planning physical facilities.

Collecting Raw Data

Many researchers are trapped by use of terminology. Terms used to express space and its use may be commonplace to some but completely foreign to faculty and even more so to the lay person. It is very important that each term be clearly defined and understood by those gathering the raw data as well as by the audience to whom the final analysis will be presented.

The development of forms to gather raw data is a two-phase problem. Each form must be designed to attain certain predetermined objectives. In addition, clear and concise instruction sheets must be developed for completing each form. This is most important to ensure the comparability of the data to be gathered.

Two forms which were very useful in the Wisconsin Study are shown in Charts I and II; these are (1) Building Data Work Sheet, and (2) Summary of Assignable and Non-Assignable Area. These are the check reins which ensure the inclusion of every room in each building as well as specifying the function of each. In addition, a single-line drawing of each building adds to the accuracy of checking the reported data.

Presenting Analyzed Data

The analyzed data can and will be used for many purposes. Primarily, it will be scrutinized by two groups, (1) the managerial personnel of the institutions and (2) state or institutional governing bodies. The interests of each are quite opposite. Managerial personnel are interested in the infinite details of utilization. On the other hand, the governing bodies are interested in the broad picture as to how the facilities are used. This is particularly true when many institutions are involved in state-wide studies.

Today, administrators and governing boards are faced with an overwhelming avalanche of written material. It has become almost impossible for them to read and digest the reams of paper crossing their desks each day. Therefore, material must be summarized according to salient points. An "overview" provides one answer to this problem. In addition, the free use of graphic presentations greatly enhances the effectiveness of a study. Bar and pie graphs seem to be particularly effective as opposed to the tabular chart or the written word. One merely has to glance at the graphic presentations to gain an insight to the presented material.

It should be noted again that clear definitions of each area analyzed are vital. The definitions should be emphasized in the printing as eye-catchers which stand out from the main portion of the text.

Perhaps the most important thing to remember is to keep the presentations and terminology simple. If the audience receiving this data cannot readily interpret it, the study is a failure. Experience shows that the person who can not interpret the data immediately assumes that someone is "pulling the wool over his eyes." Once this attitude has been created, confidence and sympathy toward the problems are lost.

The Quality Study

Quantity and Quality of space are "Siamese Twins"--inseparable!

The Wisconsin State Colleges undertook a quality study for two reasons: (1) to determine the structural quality of existing buildings and (2) to determine the adequacy of these structures in terms of their educational use.

Certainly utilization rates and future needs are affected by the quality of space. A prime example is one institution's physical education and health facilities. Structural failures have caused low utilization of these areas. Even more important are the dangers of serious injury. How was this brought to the attention of governing bodies?

A room-by-room inspection was made of (1) materials, conditions, and equipment, (2) heating, ventilating, and plumbing, and (3) electrical equipment. This was developed into a published report and integrated with colored slides. The published report, like so many reports, was read by some but more often was passed over as so much literature. However, the presentation of colored slides, coupled with verbal explanations, brought about immediate action. This approach bridged the gap between the board room and the campus. The result was a Board of Regents action to curtail physical education until structural safety tests could be made which, in turn, proved the facilities unsafe. This school is now in the stage of preliminary plans for a new physical education facility.

The Analysis Sketch*

The analysis sketch provides a realistic approach to planning and eliminates so called "blue sky" estimating. An analysis sketch may be defined as an objective method of studying the first phase of an anticipated building project. It should be noted that the term "analysis sketch" refers to a study in the planning of educational needs and is not a drawing.

The analysis sketch is designed to provide (1) a realistic estimate of cost, (2) an accurate picture of the instructional program, (3) a graphic analysis of the developed program showing areas devoted to each function and their relationship to the total program, and (4) a graphic analysis of the building site and any pertinent information relative to the site development.

As indicated previously, it is very important that all terms be clearly defined and thoroughly understood by each person participating in the collection of the data in order to ensure the success of any study. This is further emphasized by the analysis sketch which may involve architects, engineers, faculty, governing boards, and, in many instances, members of various legislative bodies. In addition, clear-cut instructions and a standard format must be developed. This is particularly true when dealing with a large program such as a state-wide system, for several projects may be in process simultaneously. In this case, it will also be necessary to have a top echelon coordinator to act as liaison between the agency and the participating architects. The analysis sketch brochure may be developed by using the following format:

1. The Program Statement. This is a statement developed by the architect and faculty working as a team. It includes a statement of the total planned instructional program and the justification of the need for the facility under study.

2. The Plot Plan. The plot plan is a graphic representation showing the location of the proposed building by a symbolic square.

3. Project Cost. This is an analysis of the total project cost.

4. Supplementary Information Concerning Project Cost. Under this heading are included the criteria for computing the various construction costs. In addition, information describing special site considerations such as pile foundations and their number and cost might be included. Further, the types and costs of special floors, walls, and other unusual areas should be included also. One of the most important items is a cost breakdown by department for general equipment. This will deal primarily with individual room requests made by the various departments involved.

5. The Graphic Analysis. The total developed program is analyzed graphically on a chart. Since the most commonly used paper size is 8-1/2"×11", this sheet must be made up of multiples of this size so as to allow this chart to become a fold-out page of the finished brochure.

6. Special Considerations. Included under this heading are items such as the following:

(a) Site consideration. Some considerations are criteria for site choice, soil test boring data, sewage problems, other utility information, and any special site development which may be necessary.

(b) Enrollments. The data should be shown by department and college for the past, present, and future. In addition, the method of obtaining the projected enrollment should be explained. For example, one method sometimes used is to take two-thirds of the past five years' average increase and project that forward five years. This is often considered quite conservative. A review of the published literature in this area will disclose the various methods which may be used in projecting enrollments.

(c) The analysis of room areas. This classification is a departmental breakdown of the total number of rooms and should include the estimated room sizes, number of student stations, the square feet per student station, and so forth. In addition, a layout is made of each proposed room indicating its dimensional size and showing the proposed installations. Also a list by rooms of the architectural, mechanical, and equipment specifications is arranged in tabular form and accompanies the layout.

7. Research. All the research done in the formulation of the project and its component parts is included in this category. The information should be infinitely complete, and all the sources of the data must be carefully described.

8. The Program. This section contains the program statement as developed by the faculty or agency involved. It includes any revisions made in the original program. From this final program statement, the computation of the cost estimate is made. Further, this section contains the names of the institution or agency members of the building committee and a project meeting memo for each meeting held during the development of the program. The meeting memo must be prepared by the architect and submitted to the project coordinator within three days after each meeting. In addition, it must be circulated to all people on the planning committee. This latter group is allowed five days to make any corrections or additions to the minutes. The meeting memo serves as a progress check on the facility being planned, which in turn is an aid to ensure the completion of the anal-

* Frederick E. Schwehr, "Eliminate Guesswork from Planning," College and University Business, May, 1961.

ysis sketch on or before the previously established completion date.

After completion of the format, data are bound into brochure-type folders. The analysis sketch is now complete and ready for submission to the examining groups for scrutiny and approval or rejection. Upon acceptance of the analysis sketch, the next phase is the development of budget preliminaries.

The inclusion of photographs greatly enhances the value of the analysis sketch. Photographs depict the quality of the existing structures and illustrate the overcrowded conditions as they may exist.

By now the reader probably is aware of the application of the analysis sketch to almost any anticipated building project. Special definitions and instructions will be needed for different kinds of buildings.

Conclusion

Effective planning of education facilities must be a team effort.

A typical higher educational planning team in Wisconsin is composed of faculty from the college concerned, an administrator from the college, a coordinator from the State Architect's office, a representative from the Board's planning and construction department, and the appointed architect.

Incidentally, this is an excellent opportunity for the administrator to "kill two birds with one stone." In the planning of the program the faculty will have an opportunity to restudy the curricula, examine its purposes and bring into sharper focus the various concepts of instruction. Thus, the docile faculty is awakened.

One of the important steps in determining the extent of future needs is to examine the quality of the existing facilities. This should not be limited to the assessment of so much brick and mortar but rather as the appropriateness of the facilities for the type of instruction now and that planned for the future. It appears this is an area where more research is needed. Understanding is needed of the differences in the quality of learning as influenced by varying physical plant environments.

Better facilities evolve when equal planning time is devoted to meeting the educational needs and the architectural design.

A "good client" must have a thorough knowledge of the educational needs justified by infinite study of the proposed program and all relevant data. In addition, it is most important to have an architect blessed with creative ability and initiative.

Perhaps the most important factor is to have a team composed of openminded individuals with a willingness to work and rework the various schemes so that the end result will be an environment which will stimulate and perpetuate higher education.

CHART 1

BUILDING DATA WORK SHEET*

Building ______________________________

Gross Sq. Ft. ______________ Net Sq. Ft. ______________

Assignable Floor Space ______________ Cubage ______________

Room No.	Length and Width	Sq. Ft. of Floor Area	Code

* Frederick E. Schwehr, "A Survey of Physical Facilities at Wisconsin State Colleges and the University of Wisconsin, Fall, 1958," (Madison: Coordinating Committee for Higher Education, Joint Staff Study XXXIII, Part I and Part II, April 1960).

CHART II

SUMMARY OF ASSIGNABLE AND NON-ASSIGNABLE AREA[2]

____
(Name of Institution)

Name of Building					Assignable Floor Space in Square Feet														Non-Assignable Space			
					Active Instructional						Active Non-Instructional							Inactive Areas				
Room Number	Department	Square Ft. of Floor Space	No. Student Stations	Student Occup.	Classroom	Teaching Laboratory	Office	Library	Auditorium	Other Instructional	Research	Extension & Public Service	Physical Education	Laboratory School	Auxiliary Enterprises	Physical Plant	Other Active Non-Instr.	Under Const. or Remodeling	Circulatory	Rest Rooms	Custodial Service	Non-Usable

[2]Frederick E. Schwehr, "A Survey of Physical Facilities at Wisconsin State Colleges and The University of Wisconsin, Fall, 1958," (Madison: Coordinating Committee for Higher Education, Joint Staff Study XXXIII, Part I and Part II, April 1960).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

SINGLE UNDERGRADUATE STUDENT COSTS AND INCOME, THE UNIVERSITY OF WISCONSIN *

L. J. LINS
The University of Wisconsin

A VERY HIGH proportion of research about students is directed toward studying levels of ability or degrees of achievement. It is also important to know something about general student characteristics, costs of college attendance, and parental backgrounds.

This study attempts to bring together information on selected student and parent characteristics, but the study deals primarily with the gathering and analyses of data concerning student and parent incomes and student costs of university attendance. Although the data are for students at The University of Wisconsin, Madison campus, and therefore should prove helpful to prospective students in anticipating costs and possible sources of income at that institution, the study also should provide a basis for comparative studies of students in other institutions.

General Procedure

Only persons were included in the study who were full-time students, who were in attendance both semesters of the 1960-61 academic year, and who were living away from their parental homes. The report is based upon a sample of about one person in every eleven of the defined population.

A questionnaire was sent to the students; over 86 percent returned usable responses. The data on parental incomes were obtained from a questionnaire sent to the parents of students who did not have foreign home addresses; 74.7 percent of the parents submitted income data.

Student Costs

The average (mean) cost of university attendance was $1,485 for students who were residents of the state of Wisconsin. Men spent $1,463 and women spent $1,530. In comparison, 10 years ago, according to the Goodnight-Trump** study, men spent $1,120, and women spent $1,150 (Table I).

From 1950-51 to 1960-61, the average cost of board and room for Wisconsin men increased by 19.5 percent and for Wisconsin women increased by 26.4 percent. If it were assumed that 1961-62 costs would reflect the $50 increase in residence halls board and room rates between 1960-61 and 1961-62, the increase from 1950-51 to 1961-62 would be 27.7 percent for men and 34.2 percent for women.

Resident fees, however, have been increased at a faster rate than board and room charges. The increase in fees from 1950-51 to 1960-61 was 83.3 percent ($120-$220) and from 1950-51 to 1961-62 was 96.7 percent ($120-$236).

Women spend considerably more for board and room and clothing than men do but less for recreation and entertainment. The average costs for books and supplies is about the same.

Non-resident men and women spend more than resident men and women spend (Table II and Chart I). The average cost, in 1960-61, for non-resident men was $2,142, for resident men was $1,463; for non-resident women was $2,409, and for resident women was $1,530. Most of the differences in the means of the amount spent by resident and non-resident students are attributable to the $380 tuition charge and the larger amounts spent by non-resident students for board and room, travel to and from home, clothing, and personal items and gifts. The tuition charge is in addition to fees paid by both resident and non-resident students. From 1960-61 to 1961-62, tuition was raised from $380 to $489 and fees were raised from $220 to $236 for the academic year.

* The data of this summary are taken primarily from: L. J. Lins, Student Expenses and Sources of Income, 1960-61 Academic Year, The University of Wisconsin, Madison Campus, Office of Institutional Studies, October 1961, pp. xi + 112.

** Scott Goodnight and Paul Trump, "Student Expenses and Financial Resources at The University of Wisconsin," Division of Student Affairs, University of Wisconsin, 1953.

Board and Room

The differences are negligible, from class to class, in the average amount spent for board and room by single undergraduate men. The average cost of board and room for single undergraduate women, however, decreases from class to class (freshman to senior); this amount is higher for women than for men in all classes. New freshman women can expect to have difficulty in locating low cost accommodations and might expect to pay more for board and room than would be necessary to pay as upperclassmen.

It is possible to have accommodations for board and room at less than $500 for the academic year. However, students who are new to the campus, particularly the women, should plan to spend at least $750 to $950 for board and room. Sixty-two percent of the freshman women paid less than $900 for board and room, and nine percent paid less than $750.

Ninety percent of the resident women spent between $510 and $1,226, a difference of $716, for board and room while 90 percent of the non-resident women spent between $517 and $1,425, a difference of $908. The same trend follows for resident and non-resident men; however, the average costs are less and the ranges are smaller.

In referring to Table III and Chart II, one notes that the highest average expenditure for board and room was by women students living in privately owned halls. The mean cost for these women was $1,164 with half of the students paying $1,252 or more, and 90 percent paying between $623 and $1,435. The fact that the mean cost in privately owned halls for women is over $267 higher than for any other living accommodation for women explains partly the high total expenditure for university attendance of non-resident women since 48.1 percent of the women students from the North Central states other than Wisconsin, 59.1 percent of the women students from the New England and Middle Atlantic states, and 38.4 percent of the women students from other states and foreign countries lived in privately owned halls.

The sample group indicates a shift from University residence halls to other types of housing after the freshman year. Whereas 63.7 percent of the freshman men and 52.0 percent of the freshman women lived in University residence halls, only 29.6 percent of the sophomore men and 44.8 percent of the sophomore women lived in University residence halls. Of the seniors, only 9.4 percent of the men and 23.9 percent of the women lived in University residence halls.

Student Income

The mean income for university attendance of single undergraduate students who were residents of the state and who were living away from their parental homes was $1,619. Although no student would have a distribution of income exactly like a distribution based upon means, the distribution is useful in obtaining a general picture of income sources. The average incomes from various sources of resident students were scholarships $82, employment during the year $154, summer employment $398, savings $169, loans $74, family contribution $660, and other income including ROTC and military reserves $82.

Family Contribution

Of their total incomes, resident men received 31.7 percent, resident women received 57.3 percent, non-resident men received 58.4 percent, and non-resident women received 80.5 percent from family contributions. Financial assistance from families is more prevalent among non-residen than among resident students; 92 percent of the male and 98 percent of the female non-resident students received some financial assistance from their families as compared with 76 percent for male and 87 percent for female resident students.

The average (mean) amounts of assistance for students receiving financial help from their families were resident men $665, resident women $1,099, non-resident men $1,482, and non-resident women $2,031. There was a wide variation, however, in the amount of assistance; for example, 90 percent of the resident men having financial help from their parents received between $96 and $1,692.

Scholarships

Of the undergraduate students in the sample, 16.6 percent of the resident men, 21.4 percent of the resident women, 16.8 percent of the non-resident men, and 5.2 percent of the non-resident women held scholarships. There were large differences in the values of the scholarships for all groups; for example, the values of 90 percent of the scholarships held by resident women were between $126 and $948.

Summer Employment and Savings

Over 69 percent of the students were employed during the summer; 74.2 percent of the men and 62.6 percent of the women were employed. Seventy-seven percent of the resident men and 70 percent of the resident women were employed during the summer.

Of the students working during the summer, 50 percent of the resident men saved over $580 for university expenses. The amounts saved from summer employment by 50 percent of the non-resident men, resident women, and non-resident women were as much as, or more than, $532, $327, and $344 respectively.

Over one-third (37.5%) had savings from sources other than summer employment. The mean savings of these persons was over $415.

Loans

Of the resident students, 6.4 percent had loans from the University, 6.8 percent had loans from relatives, and 5.0 percent had loans from other sources. At the time of the survey, 16 percent of the undergraduate students were in debt; 23.9 percent of the resident men had debts. Ninety percent of the resident men with debts had debts of between $137 and $2,360.

University Payroll and Other Work During the Academic Year

Nearly 18 percent (17.6%) of the undergraduate students worked for the University; 22.3 percent had work outside of the University. Of the resident students, 45.7 percent of the men and 42.2 percent of the women worked; of the non-resident students, 34.7 percent of the men and 20.9 percent of the women worked.

As students progress from one class to the next, apparently not only a higher proportion work but also the average number of hours worked per week increases. A higher percentage of the men than of the women in each class worked. Of the freshmen, 27.4 percent of the men and 19.2 percent of the women worked; of the sophomores, 40.9 percent of the men and 36.5 percent of the women worked; of the juniors, 47.5 percent of the men and 43.1 percent of the women worked; and of the seniors, 61.5 percent of the men and 52.2 percent of the women worked. Fifty percent of the working freshmen worked 8.5 or more hours per week as compared with 50 percent of the working seniors who worked 12.5 or more hours per week.

Half of the students who worked earned $1.06 or more per hour; half of the freshmen earned $1.00 or more per hour and half of the seniors earned $1.12 or more per hour. The average (mean) pay per hour was $.99 for freshmen and $1.38 for seniors; ninety percent of the working freshmen earned between $.59 and $1.22, and ninety percent of the seniors earned between $.90 and $3.29 per hour.

There were no great differences between the number of credits carried by working and non-working students when divided by class. The average credits carried by persons working 11 - 15 hours a week was as large as, or larger than, the average number of credits carried by students not working or working fewer hours.

Income, Occupations, and Education of Parents

Income of Parents

The average income of parents of non-resident students is higher than of parents of resident students. The percents of the parents earning $15,000 or more during 1960, divided according to the resident status of the student, were as follows: resident men, 10.7 percent; resident women, 15.3 percent; non-resident men, 38.6 percent; and non-resident women, 60.4 percent (Table IV).

Over 35 percent of the parents of resident male undergraduates earned less than $6,000 during 1960 as compared with 22.2 percent of the parents of resident female undergraduates, 11.4 percent of the parents of non-resident male undergraduates, and 3.5 percent of the parents of non-resident female undergraduates.

Occupations of Parents

Assuming that the distribution of occupations of all Wisconsin men has not changed greatly since 1950, it appears that a considerably higher proportion of the fathers of Madison campus undergraduate students is employed in professional and semi-professional and managerial and official occupations than is true of all men in the state (Table V) and that a much lower proportion of the fathers of students is in the agricultural and the skilled, semi-skilled, and unskilled occupations than is true of the total men in the state. A higher proportion of the fathers of non-resident students, in the sample, was in professional and semi-professional and managerial and official jobs than was true of the fathers of resident students in the sample.

Nearly one-third of the mothers of undergraduate students were working outside the home, that is, in occupations other than housewife. A slightly higher proportion of the mothers of non-resident students was in professional, semi-professional, and managerial and official positions than was the case of the mothers of resident students; the percentages of mothers in these occupations were resident men, 10.2 percent; resident women, 17.2 percent; non-resident men, 16.9 percent; and non-resident women, 15.2 percent.

Education of Parents

Many of the parents of undergraduate students did not attend college (Tables VI and VII). Of the fathers of resident men, 57.2 percent did not attend college and 24.5 percent had only an eighth grade education; only 27.0 percent received a baccalaureate or higher degree. Nearly 35 percent of the fathers of resident women had earned the baccalaureate degree whereas half of the fathers of non-resident students had been granted the baccalaureate degree.

A higher percentage of the mothers than of the fathers of undergraduate students completed a high school education; however, a higher percentage of the fathers than of the mothers earned a baccalaureate or higher degree. The percentages of parents with a baccalaureate or higher degree,

divided by residency and sex of the students, are as follows: male residents (mothers 17.2, fathers 27.0), female residents (mothers 23.5, fathers 34.9), male non-residents (mothers 32.7, fathers 50.0), and female non-residents (mothers 32.3, fathers 49.9).

It appears that parents who do not have a college education are more concerned that a college education be provided for their sons than for their daughters. The fact that a higher percentage of the parents of non-resident students than of resident students are college graduates is to be expected since the cost of university attendance is higher for non-resident than for resident students and since education and income are positively associated.

The average (mean) incomes of the parents of students according to the level of fathers' educations were as follows: below eighth grade, $6,523; eighth grade, $7,578; high school but no diploma, $8,081; high school diploma, $9,393; college but no degree, $12,119; baccalaureate degree, $14,605; master's, law, or equivalent degree, $19,196; and doctor's degree, $22,994. All of the income distributions are skewed to the higher incomes as shown in Table VIII. One notes that the means are higher than the medians and that there is a wide range within which 90 percent of the cases fall (fifth to ninety-fifth percentiles).

TABLE I

CHANGES IN RESIDENT SINGLE UNDERGRADUATE STUDENT COSTS MADISON CAMPUS, 1950-51 AND 1960-61

Cost Item	Men				Women			
	1960-61		1950-51		1960-61		1950-51	
	Amount	%	Amount	%	Amount	%	Amount	%
Room and Board	$ 729.20	49.9	$ 610	54.5	$ 809.01	52.9	$ 640	55.7
Books and Supplies	87.07	6.0	50	4.5	83.14	5.4	50	4.3
Travel	52.20	3.5	45	4.0	30.59	2.0	40	3.5
Clothing	76.13	5.2	70	6.2	151.04	9.9	160	13.9
Recreation and Entertainment	128.53	8.8	140	12.5	43.88	2.9	55	4.8
Other*	169.47	11.6	85	7.6	192.02	12.5	85	7.4
Fees	220.00	15.0	120	10.7	220.00	14.4	120	10.4
TOTAL	$1462.60	100.0	$1120	100.0	$1529.68	100.0	$1150	100.0

TABLE II

SINGLE UNDERGRADUATE STUDENT COSTS DIVIDED BY RESIDENT STATUS MADISON CAMPUS, 1960-61

Cost Item	Men				Women			
	Res.		Non-Res.		Res.		Non-Res.	
	Amount	%	Amount	%	Amount	%	Amount	%
Room and Board	$ 729.20	49.9	$ 802.67	37.5	$ 809.01	52.9	$1007.16	41.8
Books and Supplies	87.07	6.0	87.65	4.1	83.14	5.4	87.54	3.6
Travel	52.20	3.5	151.19	7.1	30.59	2.0	173.78	7.2
Clothing	76.13	5.2	105.77	4.9	151.04	9.9	228.69	9.5
Recreation and Entertainment	128.53	8.8	163.84	7.6	43.88	2.9	56.55	2.4
Other*	169.47	11.6	230.39	10.8	192.02	12.5	255.38	10.6
Fees and Tuition	220.00	15.0	600.00	28.0	220.00	14.4	600.00	24.9
TOTAL	$1462.60	100.0	$2141.51	100.0	$1529.68	100.0	$2409.10	100.0

*Laundry and cleaning; health; fraternity, sorority, and residence halls dues; student activities; personal items; gifts, etc.

TABLE III

COST FOR ROOM AND BOARD OF SINGLE UNDERGRADUATE STUDENTS DIVIDED BY SEX AND TYPE OF HOUSING*

Measure	University Res. Hall	Univ. Apt. or House	Sorority or Frat. House	Lodg. House Univ. Club YMCA	Privately Owned Apt.	Privately Owned Hall	Room in Private Home	Light Housekeep. Room
	Men — Type of Housing							
Mean	$779.03	$683.75	$807.09	$679.78	$741.35	$718.75	$695.78	$674.06
Median	780.90	660.00	810.00	675.00	694.00	750.00	655.71	640.00
5th %ile	708.33	332.00	577.00	483.00	481.50	--	463.00	416.00
95th %ile	845.25	870.00	1005.00	892.50	1085.50	--	1071.25	1010.00
No. of Cases	154	12	74	23	89	4	83	16
	Women							
Mean	$808.36	$737.41	$896.88	$826.67	$625.71	$1164.26	$760.00	$503.33
Median	811.92	670.00	881.25	825.00	580.00	1251.85	775.00	530.00
5th %ile	752.34	473.50	757.50	--	434.00	623.00	--	--
95th %ile	849.33	1182.50	1042.50	--	1015.00	1434.69	--	--
No. of Cases	143	27	32	3	14	89	9	3

*Does not include nine men and one woman living in rented houses or five men and five women living with relatives.

TABLE IV

INCOME OF PARENTS OF SINGLE UNDERGRADUATE STUDENTS ACCORDING TO SEX AND RESIDENCE STATUS OF STUDENT

Parental Income	Men				Women			
	Res.		Non-Res.		Res.		Non-Res.	
	No.	%	No.	%	No.	%	No.	%
Under $6,000	103	35.4	8	11.4	32	22.2	3	3.5
$ 6,000 - $ 9,999	118	40.5	18	25.7	47	32.6	11	12.8
$10,000 - $14,999	39	13.4	17	24.3	43	29.9	20	23.3
$15,000 - $19,999	9	3.1	11	15.7	10	7.0	18	20.9
$20,000 - and over	22	7.6	16	22.9	12	8.3	34	39.5
TOTAL	291	100.0	70	100.0	144	100.0	86	100.0

TABLE V

PERCENT OF FATHERS, OF SINGLE UNDERGRADUATE STUDENTS, BY OCCUPATION

Father's Occupation	Fathers of Students				Wis. Men 1950 Census
	Residents		Non-Residents		
	Men	Women	Men	Women	
	%	%	%	%	%
Professional, Semi-Prof., Managerial, Official	37.5	51.5	72.2	76.8	15.8
Clerical, Sales, Service	13.9	16.2	13.8	9.7	16.3
Agricultural	13.3	11.5	4.0	0.8	21.8
Skilled, Semi-Skilled, Unskilled	27.2	14.0	6.0	6.8	46.1
Unknown, Retired	8.1	6.8	4.0	5.9	--
TOTAL	100.0	100.0	100.0	100.0	100.0

TABLE VI

EDUCATION OF FATHERS OF SINGLE UNDERGRADUATE STUDENTS

Educational Level of Father	Residents		Non-Residents	
	Men	Women	Men	Women
	Acc. %	Acc. %	Acc. %	Acc. %
Below 8th Grade	7.3	3.2	3.0	0.8
8th Grade	24.5	16.9	5.0	3.8
High School - No Diploma	34.0	27.0	14.0	11.3
High School Diploma	57.2	47.1	32.0	24.0
College - No Degree	73.0	65.1	50.0	50.1
Bachelor's Degree	87.7	80.4	74.0	68.0
Master's, Law, and Equiv.	95.9	91.0	86.0	81.4
Doctor's (PhD., M.D.)	100.0	100.0	100.0	100.0

TABLE VII

EDUCATION OF MOTHERS OF SINGLE UNDERGRADUATE STUDENTS

Educational Level of Mother	Residents		Non-Residents	
	Men	Women	Men	Women
	Acc. %	Acc. %	Acc. %	Acc. %
Below 8th Grade	3.0	2.1	2.0	--
8th Grade	17.8	8.4	3.0	2.3
High School - No Diploma	26.8	15.2	11.2	9.8
High School Diploma	62.3	48.7	46.9	36.9
College - No Degree	82.8	76.5	67.3	67.7
Bachelor's Degree	98.1	97.4	87.7	88.7
Master's, Law, and Equiv.	100.0	100.0	100.0	97.7
Doctor's (PhD., M.D.)	--	--	--	100.0

TABLE VIII

1960 INCOME OF PARENTS OF UNDERGRADUATE STUDENTS ACCORDING TO FATHER'S EDUCATION*

Measure	Total	Father's Education							
		Below Eighth Grade	Eighth Grade	High Sch. but not Diploma	High Sch. Diploma	College but not Degree	Bach-elor's Degree	Master's, Law or Equiv. Degree	Doctor's (PhD, MD, DDM, etc.) Degree
Mean	$12,333	$ 6,523	$ 7,578	$ 8,081	$ 9,393	$12,119	$14,605	$19,196	$22,994
Median	9,000	5,350	5,800	6,800	7,100	9,550	11,150	14,000	18,000
5th %ile	3,000	1,000	1,600	3,000	2,800	5,000	4,800	7,200	6,100
95th %ile	35,000	20,000	15,000	15,000	25,000	25,000	35,000	42,000	62,000
No. of Cases	572	26	73	62	112	102	92	56	49
Percent	100.0	4.5	12.8	10.8	19.6	17.8	16.1	9.8	8.6

*Does not include 44 fathers whose occupations are unknown or who are retired or deceased.

CHART I

SINGLE UNDERGRADUATE STUDENT MEAN COSTS DIVIDED BY RESIDENT STATUS
MADISON CAMPUS, 1960-61

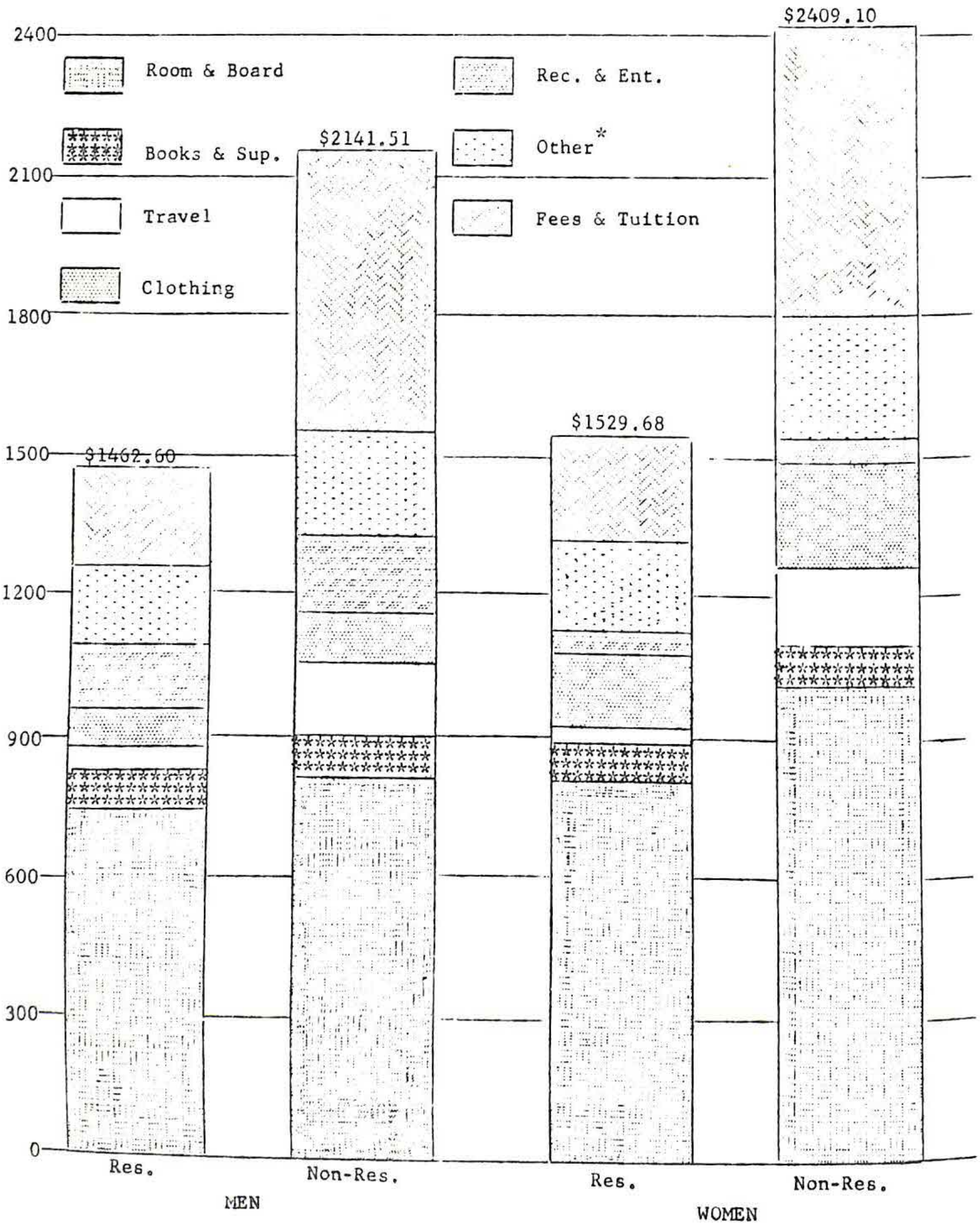
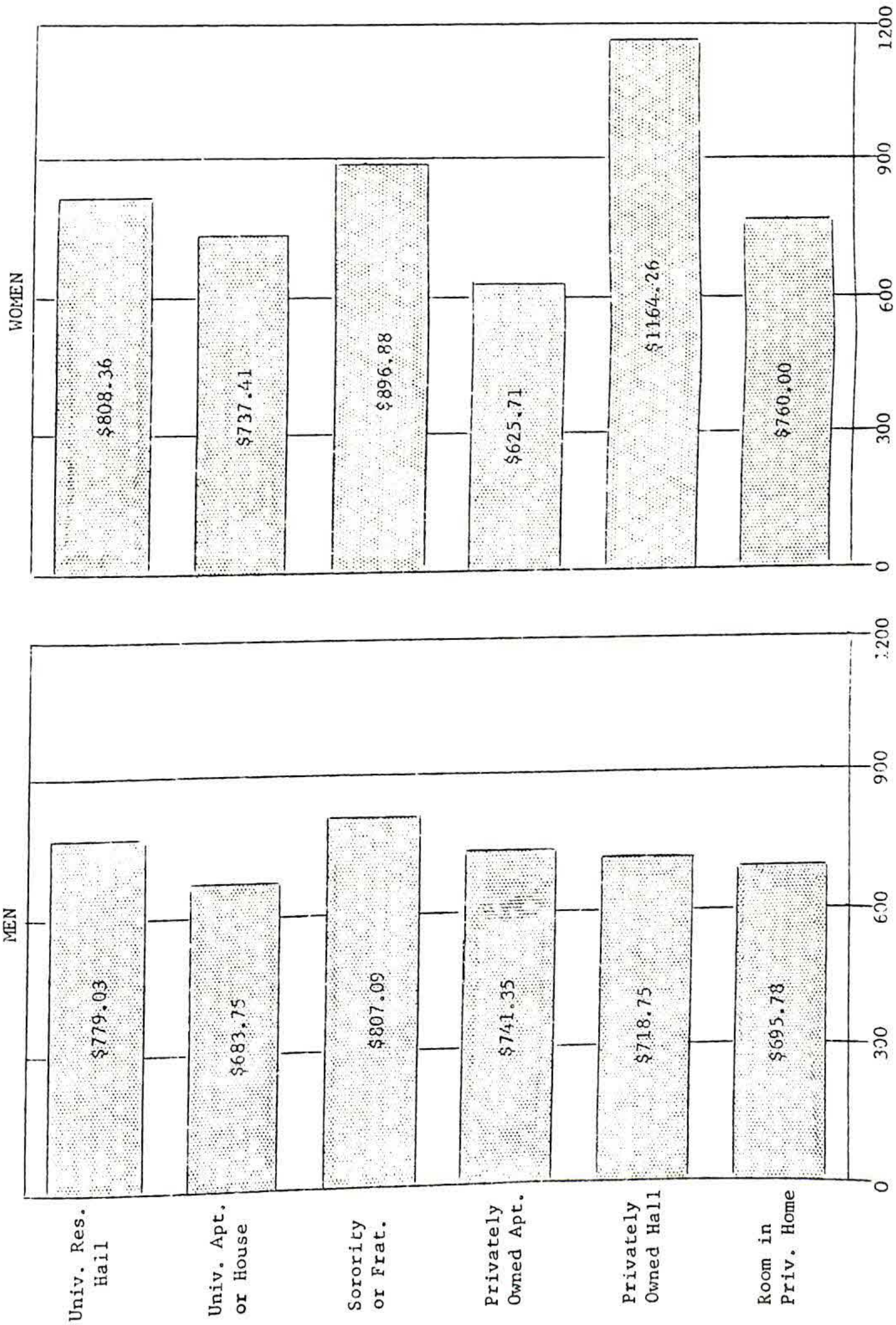


CHART II

MEAN COST OF ROOM AND BOARD OF SINGLE UNDERGRADUATE STUDENTS, MADISON CAMPUS, 1960-61



THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

APPLICATION FEES AND ROOM DEPOSITS IN SELECTED COLLEGES AND UNIVERSITIES

RALPH E. KIRKMAN
Mobile College, Alabama

Introduction

COLLEGE AND university admissions officers are confronted daily with an avalanche of inquiries from prospective students. At the March 1962 Conference on Institutional Research in Higher Education held at Northern Illinois University, one of the panelists made the observation that the admissions officer of his university would receive this year approximately 7000 inquiries from prospective students, that about 3500 students would make application for admission, that some 2200 would be accepted, but that only 1500 would actually attend the institution. This case appears to be representative of the situation which prevails in many admissions offices.

The "impending tidal wave of students" about which Ronald B. Thompson of Ohio State University wrote a few years ago is a very present reality, and prospects are for continued increases in both applications and enrollments. Students, in their desire to insure admission to a college, are in increasing numbers making multiple applications. It is not uncommon for an admissions officer to receive a mimeographed letter requesting application forms. In order to reduce the number of insincere and frivolous applications many colleges and universities require either a non-refundable application fee or an application deposit.

Purpose and Scope of the Study

The purpose of this study was to determine the extent to which state colleges and universities require a non-refundable application fee and/or a room deposit. A secondary purpose, though equally important, was to determine the range of such fees and deposits and whether or not the institutions involved were convinced that such fees and deposits accomplish intended ends.

In order to secure information from representative and primary sources, the major state university and at least one senior college in each of 49 states were requested to complete a questionnaire.* Of the 127 institutions invited to participate in the study, 117 responded.

The responses were divided into three major groups: 1) colleges and universities requiring a non-refundable application fee of all students, 2) institutions not requiring a non-refundable application fee of all students, 3) those which require some type of admission fee and/or deposit but do not require a non-refundable application fee of all students.

The Application Fee

Table I reveals the application fee and application deposit structure of the 117 colleges and universities participating in this study.

Fifteen, 23.0 percent, of the colleges and universities not requiring a non-refundable application fee or an application deposit indicated that they expect to add a non-refundable fee during the next five years. Thirty-seven, 56.9 percent, stated that they do not expect to require such a fee during the same period, while 13, 20.0 percent, reported that the matter is under active consideration. These data indicate that within five years a majority of colleges and universities will require either a non-refundable application fee or an application deposit.

Table II reports the fees charged by the institutions requiring a non-refundable application fee. The application fees assessed by most institutions are nominal and should not pose too serious a financial problem for the serious applicant.

Table III reveals the reasons given by institutions for requiring a non-refundable application fee. Twenty-five institutions indicated that the non-refundable application fee was successful in achieving its intended results. Five stated that the fee was not successful, and three failed to indicate any reaction. Of those indicating that the fee did not achieve its intended purpose, it is interesting to note that three said so because they had indicated that the fee was added to defray the cost of processing applications; two of these stated that the fee should be higher in order to take care of the costs involved, and one indicated that the state did not permit the college to retain the fee.

In general, therefore, it may be concluded that the majority of colleges and universities utilizing

*Institutions of higher learning in the State of West Virginia were not invited to participate in this study since the impetus for the study came from expected recommendations in this state. A copy of the questionnaire used in this study is contained in Appendix A.

TABLE I

THE EXTENT TO WHICH INSTITUTIONS OF HIGHER LEARNING REQUIRE A NON-REFUNDABLE APPLICATION FEE OR AN APPLICATION DEPOSIT

Designation of Fee or Deposit	Number	Percent
Colleges and universities requiring a non-refundable application fee of all students	33	28.2
Colleges and universities not requiring a non-refundable application fee	65	55.5
Colleges and universities requiring an application fee but which made refunds under certain circumstances	6	5.1
Colleges and universities requiring a deposit with each application	4	3.4
Colleges and universities requiring a non-refundable application fee of out-of-state students only	6	5.1
University requiring an application fee to be paid at the time of initial enrollment	1	0.9
University requiring an application fee from transfer students only	1	0.9
University requiring an application fee of graduate students only	1	0.9
TOTAL	117	100.0

TABLE II

APPLICATION FEE CHARGES BY 33 INSTITUTIONS OF HIGHER LEARNING

Amount of the Fee	Number Charging the Fee	Percent
$ 5.00	17	51.6
10.00	13	39.3
20.00	2	6.1
25.00	1	3.0

a non-refundable application fee are convinced that the fee serves its intended purpose.

The Room Deposit

The majority of the institutions participating in the study reported that they require a room deposit. Table IV reports the general structure of room deposits in the 109 institutions requiring such a deposit according to whether or not an admissions fee or deposit is required. The room deposits required by most institutions of higher learning are rather small although many colleges require a much more substantial room deposit than an application fee.

Summary and Conclusions

Since college and university enrollments shall continue to increase during the coming decade and since many qualified students will find it increasingly difficult to gain admission to a college or university, multiple applications will probably cause unprecedented concern to college and university admissions officers. The data reported in this study seem to warrant the following general conclusions:

1) The majority of colleges and universities will find it necessary to require a non-refundable application fee in order to defray some of the

TABLE III

REASONS GIVEN BY 33 COLLEGES FOR REQUIRING A NON-REFUNDABLE APPLICATION FEE

Reasons	Number of Institutions
Assist in the cost of processing applications	24
The fee causes a student to show a more "serious intent" in filing an application	16
Reduce the number of multiple applications	10
Discourage the unqualified from making application	2
Makes for early applications	1
Discourages "frivolous" applications	1
Stabilizes applications	1
State policy	1
To make the application more official	1
College able to make more accurate enrollment projections	1
Curtail "insurance" applications	1

TABLE IV

ROOM DEPOSITS IN 109 INSTITUTIONS OF HIGHER LEARNING

Amount of Deposit	Institutions Requiring Some Type of Admissions Fee or Deposit	Institutions Not Requiring a Non-refundable Application Fee	Institutions Requiring a Non-refundable Application Fee	Total
$ 5.00	1	5	1	7
8.00	--	1	--	1
10.00	4	10	8	22
15.00	2	4	1	7
20.00	--	5	2	7
25.00	2	23	11	36
30.00	1	1	--	2
32.00	--	1	--	1
35.00	3	--	--	3
36.00	--	1	--	1
40.00	1	--	2	3
43.00	--	1	--	1
50.00	1	3	2	6
75.00	--	1	--	1
87.00	1	--	--	1
90.00	--	1	--	1
100.00	--	--	1	1
119.50	--	1	--	1
155.00	--	1	--	1
½ all fees	--	1	--	1
$25. women, $5. men	--	1	--	1
$25. women, $10. men	--	1	--	1
$75. women, $25. men	--	1	--	1
Total fees for one semester	1	--	--	1
Amount not given	1	--	--	1
No deposit required	1	2	5	8
TOTAL	19	65	33	117

costs involved in processing applications and to reduce to a reasonable minimum the number of multiple applications.

2) In addition to a non-refundable application fee, a substantial deposit required of the student after acceptance should serve to assist colleges and universities in admitting students for whom they can provide educational opportunities.

3) Room deposits should be more substantial so as to insure the student's serious intent to attend the university.

APPENDIX A

CONCORD COLLEGE

Athens, West Virginia

1. Name of institution______________________________

2. Name of individual completing form____________________

   Position______________________________

3. Enrollment, Fall 1961-62____________________

4. Do you require a non-refundable application fee?__________ Yes or No

   If yes, please indicate the amount of the fee__________

5. Indicate reasons why you have such a fee and the major benefits derived therefrom:

   (1)______________________________

   (2)______________________________

   (3)______________________________

   (4)______________________________

   (5)______________________________

6. Are you convinced that the application fee has served its intended purpose?__________ Yes or No Comments____________________

   ____

7. If you do not have a non-refundable application fee, do you contemplate requiring one during the next five years?__________ Yes or No

8. Do you require a room deposit?__________ Yes or No If yes, amount______

9. Under what circumstances is the fee forfeited?____________________

   ____

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

WOMEN AND HIGHER EDUCATION: WITH SPECIAL REFERENCE TO THE UNIVERSITY OF WISCONSIN*

E. B. FRED
The University of Wisconsin

SOME INTERESTING FACTS CONCERNING WOMEN

1. The average American woman lives six to seven years longer than the average American man.
2. She marries younger than she did 30 to 40 years ago, and therefore has a longer portion of her life span to devote to a career.
3. The average age today of the American woman entering her first marriage is approximately 20 years—the lowest for women of the Western World in this category.
4. A generation ago nearly one-half of our college women graduates were childless at the end of their child-bearing period; but today the college women graduates show increasing fertility rates. They reproduce earlier, and they have closer spacing between the children.
5. The average woman needs approximately 15 to 20 years to raise her family.
6. On the average, she has 25 to 35 years following her child-bearing period to devote to duties other than those of the home.
7. The average age of the working woman in 1920 was 28 years; today it is 40 years. It is estimated that by 1970 one out of every three workers in the United States will be a woman.
8. Today many women who are working are occupied "at jobs that are considerably below their intellectual capacities."
9. There is great need today for more career planning for young women in college.
10. In the field of higher education, it is frequently stated that "the doors are open for women, but no one thought it important that they enter the promised land."

The Circumstances for American Women as Related to Higher Education

According to figures from the U. S. Bureau of the Census, the total number of women of all ages in the U. S. A. in 1960 exceeded the number of men by more than two-and-one-half million (2,660,000). These figures exclude persons abroad. In the near future the excess of women over men will probably reach 3,000,000 or more. In spite of the great increase in the number of women as compared with men and the removal of many of the factors that retarded college entrance, the number of women students continuing their education for higher degrees has failed to show an increase comparable to that of the men. Many of our most talented women never complete the work for higher degrees.

Among those women who do undertake higher degrees, another tendency should also be noted: the women, more than men, choose non-scientific fields of study. Figures from the U. S. Office of Education indicate that the percentage of women receiving advanced degrees in science is low. In many cases, women interested in science have not been encouraged to go on in this field. Their advisers and teachers in high schools and colleges have discouraged them—this, in spite of the fact that women in the field of scientific research have special val-

*The University of Wisconsin tables included in this report were obtained from the University of Wisconsin Coordinator of Institutional Studies, the Secretary of the Faculty, the Controller, and the Registrar. For the time and trouble these individuals have taken to supply these facts, the author wishes to express his sincere thanks.

ue. They often possess inquiring minds, imagination, and logical thinking for precise description and measurement and for critical analysis of facts and theories.

A report from the National Science Foundation[1]* on science doctorates of 1958 and 1959 combined, furnishes additional support for the statement that higher degrees are awarded predominantly to men. The NSF data are in accord with what might be expected: 91 percent of the degrees in the life sciences are granted to men; 82 percent in the behavioral sciences; and 97 percent in the physical sciences—an overall 93 percent of higher degrees in science going to men.

The figures for the number of Ph. D. degrees awarded to men as opposed to women in certain fields in the United States are given below:

Field	Men	Women
Mathematics	240	15
Physics	473	9
Biology (General)	118	29
Chemistry	960	49
Total	1791	102

The participation of women in professional occupations in the Soviet Union is totally different from that in the United States. According to Nicholas DeWitt,[2] women represent a very sizable component of the Russian professional work force. Approximately 53 percent of the Russian women are working in the various fields. The following rounded figures are given for female participation in five important work classifications:

Occupation	Approx. Percent
Physicians	75
Teachers	75
Economists and statisticians	60
Agronomists and veterinarians	39
Lawyers	32

Quite apart from the great number of women in Russia in the various professional fields, DeWitt also calls attention to the high quality of the educational program in Russia.[3]

Dr. Hoke S. Greene[4] of the University of Cincinnati calls attention to the seriousness of the American situation with the following comment: "Women scientists may determine whether the United States is to maintain scientific supremacy in the space and nuclear age."

*Footnotes will be found at end of article.

Dean Eunice C. Roberts[5] of Indiana University, at a conference on womanpower, stated, "...there is great opportunity for imaginative and daring thought in the area of women's education and employment." She concluded, quoting the late Dag Hammarskjold's judgment, "...of all the developments of the past century none may be so significant and in the long run so beneficial as the emancipation of women."

Why Should Women Be Educated?

The argument for the education of able young women was stated as follows by Sir David Smith[6] in an address to the Senate of the University of New Zealand in 1961:

> ...Universities should, I think, try to increase greatly the contribution which educated women can make to the intellectual life of the country. We should have far more qualified women on the staffs of our Universities and far more women teachers, particularly of mathematics and of science, in our post-primary schools. Somehow, it seems to me, the attitude of society needs to be changed so that able women are free at some stage in their lives, notwithstanding marriage, to engage in the deeply satisfying intellectual life of which they are capable and from which our society can derive so much benefit.

In a much lighter vein, Margaret Habein cites Daniel Defoe who, as early as 1697, gave these motives for the education of women:[7] "Women, he said, should be educated. And why? Because they then become more pleasing to men! Want of an education, said Defoe, makes a woman 'turbulent, clamorous, noisy, nasty, the devil.' On the other hand, an educated woman is all softness and sweetness, full of peace, love, wit, and delight."

T. H. Huxley (1825-95) (see H. Cyril Bibby)[8] stated:

> ...So far from imposing artificial restrictions upon the acquirement of knowledge by women, throw every facility in their way.... Let us have "sweet girl graduates" by all means. They will be none the less sweet for a little wisdom; and the "golden hair" will not curl less gracefully outside the head by reason of there being brains within.

Sir David Smith's argument[9] for the higher education of young women fits best the temper of contemporary times: "...so that able women are free at some stage in their lives, notwithstanding mar-

riage, to engage in the deeply satisfying intellectual life of which they are capable and from which our society can derive so much benefit."

In this paper, no attempt is made to review the extensive literature dealing with the education of women or to describe the many changes that have taken place in this area in recent years.[10] Without laboring such evolution it is apparent that the time has arrived when American women should carry the responsibilities for developing their minds and utilizing their intellectual capacity.

But it seems in the United States that the best use is not made of the brightest and most intellectual women. Supported by the results of numerous studies, this fact is clear.

"In this complicated and competitive world, there is no asset greater than the brain power of our people," Secretary of Health, Education and Welfare Abraham Ribicoff[11] declared recently. The desperate need for more of this vital asset is well known to all who are interested in our way of life. In spite of that need, we are utilizing only approximately one-half of our total fund of brains. Where then shall we turn? Why not, for one goal, turn to the more effective use of our wasted womanpower—the so-called "feminine fallout"?

The responsibility for better use of this great human resource lies not alone with the educator and with persons who can be instrumental in providing opportunities for women. It lies with the women themselves who must be convinced that the seeking of new goals in education will bear new satisfactions and strengths.

Dr. Ruth Hill Useem[12] of Michigan State University, declares, "Never have American women lived so long and yet been so little concerned with preparing themselves for that lengthened life. Never have we been so in need of mature, secure women and never have we had girls and women feel less secure of themselves."

In addition to this long life, women have great responsibilities to be educated whether they will be career women outside the home or homemakers and mothers. Studies have shown that on the average fathers, especially fathers of pre-school children, spend less than ten minutes per day with their children. The mother, therefore, is left alone to rely on her own resources to be both mother and father to her children and, it probably could be added, to seek her own diversions both for entertainment and intellectual stimulation while her husband is busy with business, professional, or civic activities. "That is why...that not <u>even</u> if a woman is to be married and have children <u>but especially</u> if she marries and has children, graduate education is a decided advantage."[13] Yet regardless of this and of the many other evidences of benefits resulting from education, many American women neglect their educational opportunities. Esther Raushenbush,[14] director of the Sarah Lawrence College program for the continuing education of women, writes as follows: "A great many able girls simply do not take their college education with the seriousness that corresponds to their intelligence...." She then recommends that "Positive action should be taken by colleges, especially women's colleges, in population centers, to aid the woman who did not finish her work for an A. B. degree but later in life wants to finish it... Special courses designed for such students, and agreements should be made..."

It seems clearly established that today women <u>do not have to choose</u> between marriage and a career for a happy and productive life but may combine both if the social climate is made favorable. True, the career may be limited or even abandoned during the child-bearing period, but at the close of that period, it may be expanded, provided that intellectual interests have been maintained or awakened and provided that the necessary education has been acquired or is still possible to gain.

Warren Weaver,[15] Vice President of the Alfred P. Sloan Foundation, has urged, "... as the pressure for able personnel increases, we simply must create new and appropriate opportunities for women."

A happy and productive life was the point of emphasis presented by Mary I. Bunting, President of Radcliffe College, in a discussion of the Radcliffe Institute for Independent Study. Established in November 1960, the Institute was created to help highly qualified women to carry on scholarly or creative projects. A complement to orthodox higher education is provided through various programs fashioned to suit the pattern of women's lives. Two of these programs are intended primarily for married women with family commitments.

Special Programs for the Continuing Education of Women

Recognition of a need that exists today in higher education has prompted a number of colleges and universities to provide special programs for the continuing education of women. Among these pioneering undertakings, the following should be mentioned:

1. The Radcliffe Institute for Independent Study.
2. The University of Minnesota Plan for the Continuing Education of Women.
3. The Ford Foundation Program for the Re-Training in Mathematics of College Graduate Women, Rutgers University.
4. The Sarah Lawrence College—Center for Continuing Education of Women.
5. The Michigan State University Program for Women.
6. Columbia University—the Barnard College Plan for Special Students.
7. The American Association of University Women Graduate Program of Continuing Education for Women.

These special curricula focus attention on ways that make it possible for the alumnae to resume academic work and, in many cases, for the more mature woman to complete her work for an advanced degree. For example, at Columbia University, a Barnard College alumna who has completed one year of graduate study may attend classes without paying fees.

Turner Catledge,[16] managing editor of the New York Times, writing on the responsibility to provide higher adult education in the light of modern demands, declares, "If the girls have quite a time getting educated in the first place, it is even more unforgivable that much of their education is later on permitted to grow obsolete while they are busy bringing up a family. And that's exactly where the universities' new commitment to adult education comes in—or should. Here, after all, are the institutions that ought to make it their business that a young woman who has majored in science or mathematics or history or psychology can remain up-to-date instead of becoming academically obsolescent."

The Situation for Women at the University of Wisconsin

The main objective of this report is to show what has taken place in the education of women at the University of Wisconsin during the last 20 to 40 years. Examined in the light of the preceding pages, which suggest the national picture for women, the facts bear increasing force.

For almost 100 years women have been admitted to the University of Wisconsin although they have not always been as welcome as they are today. Historians Merle Curti and Vernon Carstensen[17] state that women were first admitted to the University of Wisconsin in 1863. They state that at that time, according to James L. High, who had witnessed the coming of women, they were greeted with numerous and often bitter expressions of hostility from the men. High commented, "They came like an army with banners, conquering and to conquer; they came with bewitching curls and dimpled cheeks, and flowing robes, and all the panoply of feminine adornment; and worst of all, they came to stay."

A further reflection of early attitudes toward women in intellectual pursuits is found in John W. Hoyt's comments concerning the early days of the Wisconsin Academy of Sciences, Arts and Letters:[18] "At the preliminary meeting (of the Academy) from one-fourth to one-third of the audience consisted of women as listeners. The reason, I suppose, for this absolute silence on their part was that in Wisconsin the antagonism to women's suffrage, as a political and social measure, was at its height. Women sometimes have extraordinary common sense, expressed in extraordinary ways and this was one such occasion."

The discussion following and the tables included will demonstrate that though women "came to stay" on the Wisconsin campus, their stay still had limitations.

Comparison of Enrollment Totals for Men in Undergraduate Studies and in Graduate School with Totals for Women

Tables I and II present data in relation to the total undergraduate and graduate school enrollment of men and women for the years 1921-22 to 1960-61 inclusive. Both tables are organized to give five-year averages and not only include the numbers of men and of women but also, in the case of the women, show the percentage of the total enrollment. All of the enrollment statistics, regardless of various factors that have exerted an effect on the enrollment from year to year, are included. There has been an increase in the enrollment of both women and men; however, the proportion of women as compared with men has not shown any well defined change in the undergraduate studies and has evidenced a small decrease in the Graduate School during the 39-year period.

The variable changes in the proportion of women in the undergraduate enrollment and the decrease in the proportion of women in graduate studies may be due to a single factor or a combination of factors.

Finances—Several studies have indicated that lack of funds is a factor; it appears that when it is necessary to choose between higher education for the boy and for the girl in a family, the decision is usually made in favor of the son. When there are no funds for college, regardless of sex, the outcome is the same: waste of potential talent.

Counseling—Frequently it is failure of the parents and teachers to give young college women the essential facts required for their guidance. Many women have not been informed that the chief responsibilities in bearing children usually represent less than 15 to 20 years in a woman's life. Many do not know that the average girl graduating from college may work outside of her home for approximately 25 years or more. The desirability of gaining the most from higher education comes into focus when these facts are realized.

Marriage—Early marriage is a definite factor. Figures for 27 states show that 50.5 percent of American women marry for the first time before they are 20 years old.[19] It is within this same span that the undergraduate studies are usually undertaken and carried forward. The feeling that a college education is of little value in fitting women for the responsibilities of the home plays an important role in a girl's decision to terminate her higher education or never to begin it. How much in error these decisions are can be clearly shown by statements made by Dr. Useem and previously referred to in this paper. Moreover, although there

TABLE I

COMPARISON OF AVERAGE NUMBER OF MEN AND WOMEN, AND PERCENTAGE OF WOMEN IN ALL UNDERGRADUATE SCHOOLS AND COLLEGES UNIVERSITY OF WISCONSIN, MADISON CAMPUS*

(Average of Five-Year Periods, First Semesters 1921-22 - 1960-61)

Five-Year Interval	Average of Five-Year Interval			
	Total	Men	Women	% Women
1956-57 - 1960-61	12,734	8,190	4,544	35.7
1951-52 - 1955-56	10,479	6,809	3,670	35.0
1946-47 - 1950-51	14,440	10,383	4,057	28.1
1941-42 - 1945-46**	7,093	3,333	3,760	53.0
1936-37 - 1940-41	9,110	5,941	3,169	34.8
1931-32 - 1935-36	6,554	4,096	2,458	37.5
1926-27 - 1930-31	7,488	4,434	3,054	40.8
1921-22 - 1925-26	6,454	3,839	2,615	40.5

* Includes special students.
**Enrollment decrease due to World War II.

TABLE II

COMPARISON OF AVERAGE NUMBER OF MEN AND WOMEN, AND PERCENTAGE OF WOMEN IN THE GRADUATE SCHOOL, UNIVERSITY OF WISCONSIN MADISON CAMPUS*

(Average of Five-Year Periods, First Semesters 1921-22 - 1960-61)

Five-Year Interval	Average of Five-Year Interval			
	Total	Men	Women	% Women
1956-57 - 1960-61	3,393	2,754	639	18.8
1951-52 - 1955-56	2,678	2,220	458	17.1
1946-47 - 1950-51	2,524	2,032	492	19.5
1941-42 - 1945-46**	761	516	245	32.2
1936-37 - 1940-41	1,214	954	260	21.4
1931-32 - 1935-36	1,002	720	282	28.1
1926-27 - 1930-31	944	651	293	31.0
1921-22 - 1925-26	694	489	205	29.5

* Includes graduate students taking their work on the Milwaukee campus through 1955-56.
**Enrollment decrease due to World War II.

has been inconclusive evidence, higher educational attainment appears to be correlated to greater marriage stability.[20]

Comparison of Enrollments in Law and Medicine and of the Numbers of Ph. D. Degrees Granted for Men and for Women

The same time period and the same methods of comparison are used in the statistics offered here on enrollments and higher degrees as have been described previously for the total undergraduate and graduate school enrollments. The five-year average figures are presented in Tables III and IV. Here again, the men and women follow a familiar pattern: the total law and medical school enrollments and the total number of Ph. D.'s granted have increased, but the percentage of the Ph. D. degrees granted to women has actually decreased as has the percentage of women in law and medicine (see also the report by John B. Parrish[21]).

There is a small decrease in the percentage for women as compared with the percentage for men. The Wisconsin findings are in agreement with reports from other colleges and universities.

Home Economics Enrollment

Since the students in Home Economics are almost 100 percent women, no attempt has been made to separate the figures for men and women.

The total enrollment of undergraduate and of graduate students, as well as the overall total of Home Economics students at the University of Wisconsin, is shown in Table V. Although the figures do not go back as far in years as noted in Tables I to IV, they do give the records for 14 years.

The undergraduate enrollment in the field of Home Economics shows a decrease in the total number of women in this school and an increase in total number of graduate students. Somewhat similar figures are reported in other colleges and universities throughout the United States (see the report of Paul A. Miller[22]).

School of Education: Comparison of Number of Men with Number of Women, and the Ratio of Women to Total Enrollment

The figures for the School of Education enrollment from the years 1931-32 to 1960-61 inclusive for five-year averages are given in Table VI. In September 1951, the elementary school teaching curriculum was started at the University of Wisconsin. There was an increase in the proportion of women to men from 1946-47 to the present. For graduate students, the proportion of women during the past 15 years was less than in the five-year period 1931-32 - 1935-36; the recent proportion of women among the undergraduate group is greater than it was in 1931-32 - 1935-36. In spite of the increase in the number of women in graduate work, the proportion of women to men failed to show any increase.

College of Letters and Science: Comparison of Number of Men with Number of Women, and the Ratio of Women to Total Enrollment

At the University of Wisconsin, the College of Letters and Science has a much larger enrollment than any other school or college of the University; in fact, more than two-thirds of all the teaching in the University is carried out in this college. The figures for the total enrollment in Letters and Science and the division according to sex are presented in Tables VII and VIII.

The record extends from 1930-31 to 1961-62 inclusive and shows that the proportion of women to men in the undergraduate enrollment has changed little during the 31-year period. With the exception of the war period, the proportion of undergraduate women to men was about 40 percent. The figures for the graduate students are different (see Table VII). According to the five-year averages, the proportion of women to men has decreased.

Number of Men and Women on University of Wisconsin Faculty

From the statistics presented in Tables I to VIII, it appears that there may not be maximum use of our intellectual women. The great progress in the development of new equipment, designed to reduce labor in the home and thus provide women with more free time, has not resulted in an increase in the percentage of women going on for the higher degrees as compared with the number of men so doing; and it has not brought about a marked gain in the proportion of highly educated women as compared with men. Neither has early marriage accompanied by early retirement from child bearing apparently created an environment for increased education of women. It would be reasonable to assume that in the long run these releases would bring more and more women to an active part in college careers. But what effect have these recent changes in the estate of women had on the percentage of women on the Wisconsin faculty? The answer is given in the figures of Table IX for the years 1941-42 to 1960-61 inclusive, a 20-year period.

If the average percentage figures for the four five-year intervals are examined, it will be noted that there is a small decrease in the proportion of women on the faculty—from 23.6 percent in the first period of the 20-year span to 20.4 percent.

Instead of presenting statistics for all schools and colleges at all campuses of the University, the figures for the large basic college, the College of Letters and Science of the Madison campus, are examined next. As organized at the University of

TABLE III

COMPARISON OF AVERAGE NUMBER OF MEN AND WOMEN, AND PERCENTAGE OF WOMEN RECEIVING DOCTOR OF PHILOSOPHY DEGREES, UNIVERSITY OF WISCONSIN, MADISON CAMPUS
(Average of Five-Year Periods, Academic Years 1921-22 - 1960-61)

Five-Year Interval	Average of Five-Year Interval			
	Total	Men	Women	% Women
1956-57 - 1960-61	359	328	31	8. 6
1951-52 - 1955-56	362	331	31	8. 6
1946-47 - 1950-51	236	215	21	8. 9
1941-42 - 1945-46*	104	91	13	12. 5
1936-37 - 1940-41	155	141	14	9. 0
1931-32 - 1935-36	125	108	17	13. 6
1926-27 - 1930-31	111	99	12	10. 8
1921-22 - 1925-26	69	62	7	10. 1

*Enrollment decrease due to World War II.

TABLE IV

COMPARISON OF AVERAGE NUMBER OF MEN AND WOMEN, AND PERCENTAGE OF WOMEN IN THE LAW AND MEDICAL SCHOOLS, UNIVERSITY OF WISCONSIN MADISON CAMPUS
(Average of Five-Year Periods, First Semesters 1921-22 - 1960-61)

Five-Year Interval	Average of Five-Year Interval			
	Total	Men	Women	% Women
1956-57 - 1960-61	809	777	32	4. 0
1951-52 - 1955-56	849	812	37	4. 4
1946-47 - 1950-51	910	854	56	6. 2
1941-42 - 1945-46*	363	328	35	9. 6
1936-37 - 1940-41	686	655	31	4. 5
1931-32 - 1935-36	662	627	35	5. 3
1926-27 - 1930-31	582	546	36	6. 2
1921-22 - 1925-26	401	376	25	6. 2

*Enrollment decrease due to World War II.

TABLE V

HOME ECONOMICS ENROLLMENT AT THE UNIVERSITY OF WISCONSIN, MADISON CAMPUS RATIO TO TOTAL ENROLLMENT—FIRST SEMESTER

Year	Total			Undergraduates			Graduates		
	University	Home Economics		Total University	Home Economics		Total University	Home Economics	
	No.	No.	%	No.	No.	%	No.	No.	%
1961-62	20,118	532	2.6	14,936	469	3.1	4,396	63	1.4
1960-61	18,811	536	2.8	13,965	476	3.4	4,046	60	1.5
1959-60	17,433	534	3.1	12,932	474	3.7	3,701	60	1.6
1958-59	16,590	513	3.1	12,311	461	3.7	3,431	52	1.5
1957-58	15,929	535	3.4	12,154	495	4.1	2,990	40	1.3
1956-57	16,099	553	3.4	12,306	504	4.1	2,977	49	1.6
1955-56	15,134	543	3.6	11,649	505	4.3	2,668	38	1.4
1954-55	13,954	562	4.0	10,576	533	5.0	2,583	29	1.1
1953-54	13,346	593	4.4	10,011	562	5.6	2,525	31	1.2
1952-53	13,571	619	4.6	10,018	589	5.9	2,701	30	1.1
1951-52	14,020	670	4.8	10,139	639	6.3	2,911	31	1.1
1950-51	15,766	632	4.0	11,648	596	5.1	3,126	36	1.2
1949-50	17,690	623	3.5	13,843	601	4.3	2,822	22	0.8
1948-49	18,623	597	3.2	15,175	572	3.8	2,568	25	1.0
1947-48	18,693	624	3.3	15,658	599	3.8	2,154	25	1.2
1940-41	11,376	681	6.0	9,488	656	6.9	1,248	25	2.0

TABLE VI

COMPARISON OF AVERAGE NUMBER OF MEN AND WOMEN, AND PERCENTAGE OF WOMEN
SCHOOL OF EDUCATION
(Average of Five-Year Periods, First Semesters 1931-32 - 1960-61)

	Average of Five-Year Interval							
	Undergraduates				Graduates			
Five-Year Interval	Total	Men	Women	% Women	Total	Men	Women	% Women
1956-57 - 1960-61	1,141	295	846	74. 1	366	230	136	37. 2
1951-52 - 1955-56	1,030	317	713	69. 2	230	145	85	37. 0
1946-47 - 1950-51	1,219	551	668	54. 8	226	155	71	31. 4
1941-42 - 1945-46*	609	121	488	80. 1	87	44	43	49. 4
1936-37 - 1940-41	727	248	479	65. 9	124	83	41	33. 1
1931-32 - 1935-36	633	207	426	67. 3	92	53	39	42. 4

*Enrollment decrease due to World War II.

TABLE VII

COMPARISON OF AVERAGE NUMBER OF MEN AND WOMEN, AND PERCENTAGE OF WOMEN
COLLEGE OF LETTERS AND SCIENCE
(Average of Five-Year Periods, First Semesters 1931-32 - 1960-61)

	Average of Five-Year Interval							
	Undergraduates				Graduates			
Five-Year Interval	Total	Men	Women	% Women	Total	Men	Women	% Women
1956-57 - 1960-61	6,473	3,700	2,773	42. 8	1,881	1,490	391	20. 8
1951-52 - 1955-56	5,691	3,384	2,307	40. 5	1,428	1,141	287	20. 1
1946-47 - 1950-51	7,992	5,319	2,673	33. 4	1,534	1,190	344	22. 4
1941-42 - 1945-46*	4,319	1,701	2,618	60. 6	450	292	158	35. 1
1936-37 - 1940-41	5,702	3,600	2,102	36. 9	814	625	189	23. 2
1931-32 - 1935-36	4,298	2,608	1,690	39. 3	664	459	205	30. 9

*Enrollment decrease due to World War II.

TABLE VIII

COMPARISON OF TOTAL NUMBER OF MEN AND WOMEN IN THE COLLEGE OF LETTERS AND SCIENCE—FIRST SEMESTER*

Year	Undergraduates				Graduates			
	Total	Men	Women	% Women	Total	Men	Women	% Women
1961-62	8,962	4,996	3,966	44.3	2,616	2,023	593	22.7
1960-61	7,897	4,383	3,514	44.5	2,336	1,843	493	21.1
1959-60	6,912	3,827	3,085	44.6	2,080	1,637	443	21.3
1958-59	6,116	3,523	2,593	42.4	1,839	1,465	374	20.3
1957-58	5,622	3,246	2,376	42.3	1,640	1,302	338	20.6
1956-57	5,820	3,523	2,297	39.5	1,512	1,205	307	20.3
1955-56	6,006	3,564	2,442	40.7	1,360	1,082	278	20.4
1954-55	5,760	3,399	2,361	40.9	1,296	1,007	289	22.3
1953-54	5,535	3,200	2,335	42.2	1,359	1,091	268	19.7
1952-53	5,530	3,311	2,219	40.1	1,469	1,186	283	19.3
1951-52	5,624	3,445	2,179	38.7	1,655	1,337	318	19.2
1950-51	6,361	4,086	2,275	35.8	1,828	1,478	350	19.1
1949-50	7,453	5,049	2,404	32.3	1,710	1,368	342	20.0
1948-49	8,174	5,557	2,617	32.0	1,577	1,206	371	23.5
1947-48	8,657	5,795	2,862	33.1	1,360	1,025	335	24.6
1946-47	9,315	6,107	3,208	34.4	1,197	875	322	26.9
1945-46	4,953	1,374	3,579	72.3	536	298	238	44.4
1944-45	3,759	801	2,958	78.7	306	149	157	51.3
1943-44	2,974	721	2,253	75.8	308	179	129	41.9
1942-43	4,654	2,529	2,125	45.7	419	292	127	30.3
1941-42	5,256	3,081	2,175	41.4	680	542	138	20.3
1940-41	5,760	3,560	2,200	38.2	815	650	165	20.2
1939-40	5,710	3,580	2,130	37.3	823	642	181	22.0
1938-39	5,850	3,724	2,126	36.3	859	657	202	23.5
1937-38	5,815	3,738	2,077	35.7	780	594	186	23.8
1936-37	5,373	3,398	1,975	36.8	793	580	213	26.9
1935-36	4,942	3,118	1,824	36.9	654	472	182	27.8
1934-35	4,321	2,714	1,607	37.2	583	403	180	30.9
1933-34	3,818	2,266	1,552	40.6	562	395	167	29.7
1932-33	3,918	2,329	1,589	40.6	721	487	234	32.5
1931-32	4,489	2,613	1,876	41.8	800	539	261	32.6
1930-31	5,317	3,019	2,298	43.2	787	536	251	31.9

*Graduate students taking their work at the Milwaukee Extension Center during 1955-56 and prior years are included. Graduate students taking their work at the University of Wisconsin-Milwaukee for the years 1956-57 and succeeding years are excluded. Includes nursing for 1955-56 and prior years; pharmacy for 1949-50 and prior years; and commerce for 1944-45 and prior years.

Wisconsin, this college provides approximately 68 percent of all of the instruction at the University. The numbers of men in the ranks of professors and instructors as contrasted with the numbers of women are shown in Tables X, XI, and XII, for a period of seven years from 1954-55 to 1960-61. (In all three of these tables the administrative officers and the non-teaching members of the staff have been omitted.) The number of women in the lower ranks, assistant professors and instructors, show a slight increase (Table X) while the number of women in higher ranks show a slight decrease. The percentage figures for both groups are small.

Table XI gives the figures for the two groups (men and women) in the higher ranks (professors and associate professors) from 1954-55 to 1960-61 inclusive. In 1954, women represented 5.4 percent of the professors and associate professors, and in 1960-61 the women on the faculty represented 5.0 percent of the total faculty in these ranks, a slight drop in the percentage of women. However, this 5.0 percent in 1960-61 is higher than the percentages of all years intervening between 1954-55 and 1960-61.

Perhaps it is noteworthy at this point to say that according to a report from the Woodrow Wilson Foundation,[23] "...the proportion of women on liberal arts college faculties is about 14 percent..." At The University of Wisconsin in the College of Letters and Science the proportion is less than 10 percent.

A recent report by John B. Parrish[24] is of special interest. Parrish carried out a study of the status of women on faculties engaged in top level teaching and research. Of the 18 institutions included in his report, he found that women represented approximately seven to 10 percent of the faculty. The percentage figures were much higher in the case of instructors and assistant professors. As a result of his study, he concluded, "Failure of women to advance up the academic ladder appears rather striking in view of their substantial contributions at lower rank levels."

Comparison of Major Fields of Interest of Men and Women—Social Sciences, Humanities, and Natural Sciences

In order to obtain some idea of the variation in fields of interest of the men and women students at Wisconsin, 40 of the top ranking men students and 40 of the top ranking women students receiving the Bachelor's degrees for 1960-61, as shown by high grade-point averages for their four years in college, were selected and their major fields recorded:

	Men	Women
1. Humanities	8	20
2. Social Sciences	8	14
3. Natural Sciences	24	6

The results as summarized bring out strikingly the great differences between men and women students as related to the three major divisions of studies. The major numbers of top ranking men are interested in the natural sciences; and the ratio of men to women in this field is four to one. The women, on the other hand, are primarily interested in the humanities and the social sciences. Perhaps the sample for this test is too small to be of great accuracy, but at least it gives an indication of the differences in the fields of interest of these two high grade-point groups divided by sex.

Summary

The plain facts concerning enrollment of women at The University of Wisconsin are these:

The proportion of women to men in the total enrollment figures for the undergraduate schools, with the exception of home economics, has shown little change for 31 years. The figures for the women are 33.3 percent for 1960 and 34.4 percent for 1961. They are somewhat lower than the percentage of women in the total U. S. enrollment figures for full- and part-time students. According to the U. S. Office of Education, women made up 37 percent of the 1960 enrollment at institutions of higher learning and 37.7 percent in 1961.

The figures for Wisconsin's Graduate School point more sharply to a wasted resource: the proportion of women going on for the Ph. D., as compared with men seeking the Ph. D., has decreased. The case might be stated this way: among people who have the ability to pursue college studies, men show a higher proportion in college enrollment, and this is especially true in enrollment in graduate studies.

These questions are worthy of repeating again: Why is there so much unused brain power among women—those who never go to college, or who, if they do go, drop out? (The drop-out rate for women is greater than that for men.) Is it because of early marriage? Is it because of lack of finances? Is it because certain individuals have an idea that it is "unladylike" to take a higher degree and especially to follow a career in science? Is it because women have not been informed of the great improving influence which higher education can exert upon their lives today and tomorrow? Is there lack of incentive; if so, what incentive is needed?

From all sides complaints are heard about the fact that women are a neglected resource in the development of highly trained persons for our expanding economy and technology. But those who complain frequently do not suggest a program which is able to tap this resource.

Recruitment

Larger numbers of women cannot be employed

TABLE IX

COMPARISON OF THE PERCENTAGE OF WOMEN ON THE FACULTY OF THE UNIVERSITY OF WISCONSIN ALL CAMPUSES 1941-42 - 1960-61*

Year	Percent Women	Five-Year Average
1960-61	19.6	
1959-60	19.7	
1958-59	20.4	20.4
1957-58	21.5	
1956-57**	20.7	
1955-56	18.0	
1954-55	18.6	
1953-54	19.1	19.1
1952-53	19.8	
1951-52	20.0	
1950-51	20.0	
1949-50	20.9	
1948-49	22.6	22.2
1947-48	23.0	
1946-47	24.6	
1945-46	24.6	
1944-45	25.7	
1943-44	24.9	23.6
1942-43	22.8	
1941-42	20.1	

* All branches of the University (instructors, assistant professors, associate professors, and professors) excluding administrators.

**Milwaukee Extension Center merged with Milwaukee State Teachers College to form University of Wisconsin-Milwaukee.

TABLE X

COMPARISON OF TOTAL NUMBERS OF MEN AND OF WOMEN ASSISTANT PROFESSORS AND INSTRUCTORS IN THE COLLEGE OF LETTERS AND SCIENCE, UNIVERSITY OF WISCONSIN, MADISON CAMPUS 1954-55 - 1960-61

Year	Total	Men	Women	% Women
1960-61	165	137	28	17.0
1959-60	143	120	23	16.1
1958-59	137	114	23	16.8
1957-58	122	99	23	18.9
1956-57	120	102	18	15.0
1955-56	118	104	14	11.9
1954-55	111	95	16	14.4

TABLE XI

COMPARISON OF TOTAL NUMBERS OF MEN AND OF WOMEN FULL AND ASSOCIATE PROFESSORS IN THE COLLEGE OF LETTERS AND SCIENCE, UNIVERSITY OF WISCONSIN, MADISON CAMPUS
1954-55 - 1960-61

Year	Total	Men	Women	% Women
1960-61	298	283	15	5.0
1959-60	279	267	12	4.3
1958-59	262	250	12	4.6
1957-58	243	233	10	4.1
1956-57	236	227	9	3.8
1955-56	241	230	11	4.6
1954-55	224	212	12	5.4

TABLE XII

COMPARISON OF TOTAL NUMBERS OF MEN AND OF WOMEN PROFESSORS (ALL RANKS) AND INSTRUCTORS IN THE COLLEGE OF LETTERS AND SCIENCE, UNIVERSITY OF WISCONSIN, MADISON CAMPUS
1954-55 - 1960-61

Year	Total	Men	Women	% Women
1960-61	463	420	43	9.3
1959-60	422	387	35	8.3
1958-59	399	364	35	8.8
1957-58	365	332	33	9.0
1956-57	356	329	27	7.6
1955-56	359	334	25	7.0
1954-55	335	307	28	8.4

in occupations which require advanced and highly specialized education unless larger numbers of women present themselves for specialized training. Despite demonstration of the capacity of women to distinguish themselves in science, medicine, and practically any field of human enterprise, few women in our culture choose the paths which lead to these careers. It is doubtful that this situation will change except through strong shifts in our cultural milieu or in response to urgencies dictated by clear and present national or international needs and spelled out in well-defined programs designed especially for the development and use of the Nation's womanpower.

Utilization

Women are understandably reluctant to enter programs of long and expensive training if their opportunities for service are either limited in number or less rewarding in type of service or compensation than opportunities for men who have similar qualifications.

Currently, employers in such fields as science, engineering, and medicine, including academic departments of universities, tend to choose the "blue ribbon" man instead of the "blue ribbon" woman. Some observers suspect that a "second place" man would win over the "blue ribbon" woman. Consideration of factors other than individual merit is reasonable in some instances, but sometimes these considerations have doubtful validity.

If, however, the number of highly qualified women applicants were equal to the number of highly qualified men applicants, the apparent discrimination might not persist. Certainly the likelihood of women being selected would be much greater. The sheer number of highly qualified women applicants would command the attention of employers and operate against neglecting the resource they represent.

Therefore, if the needs of the Nation so dictate, and if womanpower is a present and ready resource for recruitment, development, and utilization to meet those needs, a program designed specifically to claim this resource should be established. In the author's opinion, such a program must be sponsored at the national level if it is to provide the necessary incentive. It should be a program designed specifically for women and keyed to their special needs and potentialities. Opportunities for research, teaching, and productive contribution in fields now relatively neglected might be emphasized.

Such a program might illustrate a concept of the power and dignity of a womanhood fulfilled in a free society dedicated to human advancement—a concept in sharp contrast to the circumstances in some countries where women are highly trained—yes—but to be used as economic units in a slave society.

Meanwhile, each institution might consider whether its programs for counseling, advising, scholarships, fellowships, and other incentive projects, as well as faculty recruitment and utilization programs, are future-facing with respect to the prospective role of women in our society. Most of the practices in these programs are predicated upon principles of equal treatment of sexes in open competition.

The writer believes there must be greater attention to the separate and specific goal of providing incentive for more girls to continue their schooling to advanced levels and to employing more women in the teaching and research staffs of more university and college departments as soon as these women reach the market.

FOOTNOTES

1. The Science Doctorates of 1958 and 1959, National Science Foundation, NSF 60-60, 1961, p. 15.
2. Nicholas DeWitt. Education and Professional Employment in the U. S. S. R., National Science Foundation, 1961, NSF 61-40, pp. 492-93.
3. Ibid., p. 29.
4. Hoke S. Greene. "Women Remain Greatest Untapped Source of Scientific Talent," Higher Education in New England, New England Board of Higher Education, Vol. V, No. 2 (June 1961), p. 7.
5. Midwestern Conference on Womanpower, sponsored by Indiana University, May 21-22 (1958), p. 23.
6. David Smith. The Recent Past - Agricultural Education - International Education Communication - University Education for Three Classes of Persons - The End of the University of New Zealand. An address to the Senate of the University of New Zealand at its meeting at the Victoria University of Wellington, Wednesday, August 23, 1961, p. 21.
7. Margaret Habein. The Education of Women—Signs for the Future, (Washington, D.C.: American Council on Education). A report of a Conference on the Present Status and Prospective Trends of Research on the Education of Women, held at Rye, New York, October 27-30, 1957, p. 97, citing Daniel Defoe, "The Academy for Women," Essay upon Projects (London: Cassell and Co., Ltd., 1894), pp. 164, 175.
8. H. Cyril Bibby. T. H. Huxley, Scientist, Humanist and Educator (New York: Horizon Press, 1960), p. 36.
9. Smith, op. cit., p. 21.
10. Habein, op. cit., pp. 1-153.
11. Abraham Ribicoff. Address before the American Council on Education, Mayflower Hotel,

Washington, D.C., October 5, 1961, pp. 15-16.

12. Ruth Hill Useem. "Changing Cultural Concepts in Women's Lives," Journal of National Association of Women Deans and Counselors, XXIV (October 1960), p. 29.
13. "Where Are the Women," in The Michigan State University Magazine, March 1962, p. 17. Based on a talk given by Dr. Ruth Hill Useem.
14. Esther Raushenbush. "Unfinished Business: Continuing Education for Women," The Educational Record, XLII (October 1961), pp. 267-68.
15. Warren Weaver. "A Great Age for Science," Goals for Americans, The American Assemby, Columbia University, 1960, p. 116.
16. Turner Catledge. "The Responsibilities of Colleges and Universities to Provide Higher Adult Education in Terms of Today's Demands," Conference on Higher Adult Education in the South, Proceedings, a report of sessions held February 22-24, 1961, at Gatlinburg, Tennessee, p. 28.
17. Merle Curti and Vernon Carstensen. The University of Wisconsin—A History, 1848 to 1925, Volume I, Madison, Wisconsin, 1949, p. 193.
18. A. W. Schorger. A History of the Wisconsin Academy of Sciences, Arts, and Letters, transactions of the Wisconsin Academy of Sciences, Arts, and Letters, ca. p. 1. (Manuscript to be published about June 1963.)
19. U. S. Department of Health, Education and Welfare, Public Health Service, National Office of Vital Statistics. The Vital Statistics of the U. S. 1959, Volume I, Table 6, pp. 59-60.
20. Paul C. Glick and Hugh Carter. "Marriage Patterns and Educational Level," American Sociological Review, XXIII (June 1958), p. 297. See also Ray E. Baker, Marriage and the Family (New York: McGraw-Hill, 1953), pp. 468-69. "Women's Stake in the College Price Tag," Joint Office of Institutional Research, 1785 Massachusetts Avenue, NW, Washington, D.C., 1961, p. 3. Mabel Newcomer, "Women's Education: Facts, Findings, and Apparent Trends," Journal of National Association of Women Deans and Counselors, XXIV (October 1960), p. 38.
21. John B. Parrish. "Professional Womanpower as a National Resource," Quarterly Review of Economics and Business, I (February 1961), pp. 54-63.
22. Paul A. Miller. "Higher Education in Home Economics: An Appraisal and a Challenge," Proceedings, American Association of Land-Grant Colleges and State Universities, 74th Annual Convention, 1960, pp. 244-46.
23. Woodrow Wilson National Foundation. Report of Activities, 1957-61, Princeton, New Jersey, October 1961, p. iv-43.
24. John B. Parrish. "Women in Top Level Teaching and Research," American Association of University Women Journal, LV (January 1962), pp. 99-103; 106-107.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

NON-IMMIGRANT FOREIGN STUDENTS: A SURVEY OF THEIR NEEDS AND INTERESTS

JOSEPH TANENHAUS and SIDNEY G. ROTH
New York University

IN THE academic year 1960-1961, New York University enrolled some 3200 students who were foreign citizens — a larger number perhaps than any other institution of higher education in the United States. Nearly two-thirds of these persons (1944) held temporary visas.[1*] Concern about the educational and related needs and interests of these foreign students on temporary visas recently led New York University to ask their cooperation in a study of their problems, with a view toward improving the services the University offers in helping to resolve them. Some of the data developed in this study seem to be of general interest and will be summarized in this article. As a preface to this summary, a brief discussion of the study methodology follows.

On May 1, 1961, a twenty-page questionnaire was mailed to the 1944 persons holding temporary visas and who were in residence for at least one of the two regular semesters during 1960-1961. The questionnaire was designed by Dr. John Dale Russell (then Director of NYU's Office of Institutional Research), under the general guidance of an ad hoc Committee on Foreign Students. In addition to the usual personal data, the questionnaire contained clusters of items about academic, physical, financial, social, and cultural needs and interests of the respondents. The last question solicited comments about subjects not fully developed. In answer to this invitation for comment, fully a third of the respondents (384 by the time data processing got under way), appended statements, many exceeding 200 words in length. In addition, literally hundreds of explanatory notes were scattered throughout the completed questionnaires. These qualitative data have been found to be invaluable in interpreting the quantitative responses.

When compared with known parameters such as sex and country of citizenship, the respondents seemed to constitute a fairly good sample not only of New York University's non-immigrant foreign student population but also of all foreign students on temporary visas in the United States. Although differences in questionnaire construction make direct comparisons hazardous between NYU's study and the excellent investigation directed by Dr. Homer Higbee, we quite agree with his judgment that our "findings are thoroughly consonant with (his) in those areas where we were looking for approximately the same things."[2] Unfortunately, the small number of African students in our sample makes any comparison questionable between our data and the <u>IIE Survey of the African Student</u>.[3]

Academic Problems

The central interest of any institution of higher education is, of course, the academic needs of its students. And certainly primary among the academic objectives of the foreign student is acquiring the knowledge and skills requisite for pursuing the career of his choice. In fact, what was most impressive about the answers to a question concerning the uses to which respondents planned to put their education once they got back home, was that only 5.5 percent selected "no definite purpose in mind."

It seems generally clear that foreign students have as their most important objective the acquisition of knowledge and skill. However, they tend also to have a secondary objective which is not always compatible with the first: obtaining some tangible certification of their American educational experience. More than three-quarters of the sample felt that an American degree was either essential (46.6%) or "highly desirable, but not absolutely essential" (31.3%). Another 9.1 percent felt they could accept certificates or diplomas in lieu of actual degrees. A mere 6.0 percent expressed little interest in official certification of any sort.

By far the most important service an institution of higher education provides to foreign students is the academic program it offers. The respondents to the survey underscored their awareness that the courses offered tend to be heavily oriented towards the needs of domestic clientele. To be sure several courses in English have been

* Footnotes will be found at the end of the article.

designed especially for foreign students. Some measure of the success of these courses, required of all students who do not exhibit the necessary proficiency in English usage, was revealed when only 5.7 percent of those who have been in residence for at least 12 months found their lack of familiarity with the English language a serious handicap in their studies. In fact, some felt the University program in English unnecessarily demanding. As one respondent put it, the emphasis placed on proper English usage is "to(sic) much."

It must be recognized, however, that something more might be attempted in other fields. In answer to the question, "If the University offered a special course designed to help you apply what you have learned to the situation in your home country, would you register for such a course?", nearly 75 percent said yes. It was comforting to learn that only a handful (2.6%) thought few or none of the courses they had taken would be useful when they returned home.

That a student's primary objective in obtaining knowledge and skill and his secondary objective in receiving official certification from an American university are not always complementary is supported by the reactions to the question, "Do you feel that requirements for a University degree are preventing you from taking courses you really need?". Nearly one-fifth (19.0%) replied that degree requirements were a more or less serious obstacle to taking courses they really needed. Answers to related questions and appended glosses suggest that much of the unhappiness exists because these students are not granted credit for what they deem to be comparable work taken at home. Quite a number of persons entered earnest pleas for some more satisfactory method of evaluating the courses they took before coming to the United States. These pleas are well taken. However thorny the thicket which crowds the path, a better way must be found for evaluating the credentials of students from abroad.

If a university's academic program is the most important service it can offer, then the guidance provided for using it best requires serious attention. The respondents made it very clear that they preferred to get advice on academic problems from their course instructors and not from full-time counselors, the Foreign Student Center staff, or other administrative officials. In fact, foreign students seemed to show much less hesitation in approaching their instructors for advice on academic matters than their American peers. Less than 20 percent (18.0%) of the respondents had never asked for it. And those who did seek the counsel of their instructors were generally pleased with the assistance they received. Only one respondent reported his instructors hostile, and but seven others (1.8%) thought them generally little interested in the respondents' academic problems. More than two-thirds of the sample, on the other hand, found the instructors they approached generally cordial and helpful.

Personal Problems

The line between academic and personal needs is not always sharply defined. Economic problems are an illustration of this. Fully half of the respondents (50.5%) said that they had experienced financial difficulties since coming to NYU. Of this group, 47.4 percent indicated that a major reason for their discomfiture was that expenses were higher than they had anticipated. Attached comments leave a firm impression that respondents considered the University at least partially to blame because satisfactory information about expenses was not made available before they left for the United States. (It should be noted parenthetically that some persons also felt information about the educational programs of American universities should be made much more easily accessible in their homelands.) In fact, some students in the sample believed their economic problems so much the University's responsibility that they felt it obliged to help mitigate them by such methods as providing employment, remitting tuition for all foreign students, waiving course requirements for those who were pressed for funds, and securing exemptions from sales and income taxes.

Whatever one may think of a university's obligation to underwrite certain of these financial needs (or is it the Federal Government's?), there are certainly two things that an institution of higher learning can readily do. One is to make accessible to students planning to come to the United States more adequate data about the financial commitments entailed in attending school here. Second, since one student in twenty is financially embarrassed from time to time by delays in the transmittal of funds from home, a university can establish a short-term revolving loan fund to alleviate this hardship.

Some of the other personal needs had substantial economic overtones. For example, 72.9 percent of the 140 respondents who experienced difficulty in obtaining suitable housing cited rental costs as the principal reason. Again, 22.7 percent of the 154 in the sample who found themselves unable to take satisfactory advantage of New York City's cultural facilities pointed to insufficient funds for admission fees as the major reason.

Most of the respondents expressed some concern about their difficulties in getting to meet other persons. Only 39.3 percent seemed entirely satisfied with the social contacts they had made. The most serious obstacle to developing a satisfactory social life was the slender amount of time left free from studies. Still, 15.1 percent reported that the dearth of opportunities to meet others was at fault. This lack of opportunity

proved to be very much a function of length of time spent at the University. More than one respondent in three who had been here for less than six months (36.4%) complained about the absence of opportunities to meet other persons, but the percentage dropped sharply for the group which had been at the University for at least six months. Nevertheless, as many students pointed out in written comments, getting to know other students is an important part of the educational experience which a university can rather readily do much to facilitate.

Faculty assistance, which turned out to be so useful in dealing with the foreign student's academic difficulties, did not extend to the handling of his personal needs. It is not so much that instructors proved hostile to the foreign students who approached them with concerns of a non-academic nature, for this tended not to be the case. Only three respondents indicated that they had actually been rebuffed, and 18 others (4.7%) that instructors were only occasionally willing to help them work out personal affairs. The reason why faculty members proved to be of small assistance in helping foreign students cope with their non-academic troubles is that most students (66.7%) never approached them on these matters. That many students would, if they felt free to impose, is strongly suggested by the answers to a variety of other questions.

Non-Student Roles

The persons in the sample play several roles in addition to that of student. Among these are tourist and unofficial amabassador from abroad. When the respondents return to their homelands, they will assume, among others, the roles of university alumnus and interpreter of the United States. The questionnaire was designed to develop data primarily relating to the foreign students' role within the University, and this report has concentrated on it. However, several items have some relevance for analyzing their roles as ambassador from abroad and spokesman about the United States when they return home.

The respondents have a lively appreciation of their opportunity to play the role of unofficial ambassador. Only one person in six (16.7%) did not consider ft his responsibility to speak on behalf of his country. Yet even of this small group 70 percent had done so to some extent anyway. The most obvious opportunity for the foreign student to explain his country to Americans is through giving more or less formal talks about it. Fully half of the sample (52.3%) had given at least one such speech or talk. Moreover, 25.8 percent had done so four or more times. While a few persons (2.6%) reported that they have been called upon more than they would like, nearly half of the sample (45.8%) expressed the desire for additional opportunities to present talks about their homelands.

The respondents may not have regarded it as such, but another kind of opportunity for explaining their countries to American citizens is presented when the foreign student is invited as a guest to a citizen's home. Opportunities of this sort were far more frequent than initially expected. Nearly nine-tenths of the sample (87.2%) had been guests in homes of United States citizens at least once. This high incidence of home entertainment is to some extent a result of the program whereby the Columbia Broadcasting System cooperates with New York University in arranging a home invitation on Thanksgiving Day for every foreign student who is prepared to accept one. But this cannot be the major explanation because many of the foreign students had been home guests on numerous occasions — 60.7 percent from four to 20 or more times. Although some respondents felt that they have been entertained more from a sense of duty than a desire for friendship, these were relatively few in number (4.4%).

The role foreign students will play in the future as interpreters of the United States to their fellow countrymen is much more difficult to assess. If the responses are representative, Americans can be rather sanguine about the manner in which our country will be portrayed when the foreign student returns home. Less than one person in five felt that he was not getting much of an opportunity to develop an understanding of American life and institutions; 41.1 percent thought that their opportunities had been thoroughly adequate. Another 40.6 percent did not consider them unsatisfactory but would like to have seen them increased.

There was no general agreement as to what sort of experiences would improve the foreign student's understanding of American life. But enough students (32.0%) indicated an interest in taking a course covering broadly all aspects of American life and institutions to justify such an offering.

Conclusions

This report has been so interspersed with conclusions that only one additional comment seems appropriate. Much of the data developed have not been included in this study because they are applicable only to New York University. These unreported data have proved valuable enough to NYU so that it is believed that other institutions with large populations of foreign students will find it highly worthwhile to undertake surveys of their own.

FOOTNOTES

1. The figures New York University reported to the Institute of International Education (Open Doors, 1961, pp. 11, 44), differ from those reported in the text primarily because the Institute does not include in its surveys the

foreign students on temporary visas enrolled in NYU's non-degree granting Division of General Education.

2. Letter of November 10, 1961. The Higbee report is The Status of Foreign Student Advising in the United States Universities and Colleges, East Lansing, Michigan: 1961.

3. James M. Davis, Russell G. Hanson, and Duane R. Burnor, IIE Survey of the African Student: His Achievements and His Problems, New York: 1961.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

COLLEGE FRESHMEN'S MOTIVES FOR GOING TO COLLEGE AND ACADEMIC ACHIEVEMENT

GERHARD LANG, AMEDEO G. SFERRA and ANN K. KNUDSEN
Fairleigh Dickinson University

INSTITUTIONS OF higher learning are confronted with the task of accommodating qualified students who, in increasing numbers, wish to be admitted (7, 11). This task is quite formidable in view of the difficulty to select, fairly and efficiently, from among many applicants those who seem best qualified to profit from advanced training (2). Moreover, it is noted that the attrition rate of college students is distressingly high (6) and that many students who remain in college are underachieving (5, 12). Research in the area of college admission traditionally has focused on relating intellective predictors to intellective criteria but has recently included the consideration of non-intellective predictors and their relationship to intellective and/or non-intellective criteria (2). For instance, student perceptions of college (14), motivation for going and remaining in college (3, 10), and psychological needs of college freshmen (8) are some of the factors which significantly predict academic achievement and/or persistence in college (13).

The purpose of this study was to determine the extent to which qualitative differences in freshmen's motives for going to college are related to their academic achievement, based on instructors' estimates.

Subjects and Instrument Used

The sample used in this study consisted of 38 male and 49 female college freshmen, 17-19 years of age, who entered Fairleigh Dickinson University in the fall of 1959. These students represented a cross-section of the four major colleges (Liberal Arts, Science and Engineering, Business Administration, and Education).

Students' motives for attending college were elicited by a scale which was especially developed for this study, the L-M Scale of Motives for Going to College, hereafter referred to as "the scale." The construction of the scale and evidence regarding validity and reliability are discussed elsewhere (9). The scale consists of 27 reasons, reflecting varying degrees of social desirability, which might prompt a high-school senior to pursue a college education. These reasons, listed in Table I, comprise three types: "occupational-economic" (nine reasons), "academic-idealistic" (seven reasons), and "conformity-social" (11 reasons).

Procedure

The scale was administered to 67 male and 69 female freshmen while enrolled in orientation classes. The students were directed to read through the list of 27 reasons and then to check three reasons which have influenced them most in going to college, three reasons which have influenced them least, six reasons which influenced them much, and, finally, six reasons which they regarded as slightly influential. In the analysis of data, only most and least influential reasons were used. On the basis of students' responses, seven motive categories were established which are enumerated in Table II.

In many previous studies, students' achievement levels were determined by noting the discrepancy, if any, between intelligence test scores (representing predicted achievement) and grade-point ratios (constituting the actual achievement). The validity of this procedure has been severely questioned (1). It was felt, therefore, that instructor ratings, obtained for the present study, would give a more meaningful and direct estimate of students' academic achievement. Toward the end of the 1959-60 academic year, 81 instructors in the various disciplines were asked to rate, on a five-point scale, each student's academic achievement with respect to his intellectual capacity. The analysis of data is based on only those subjects for whom five faculty ratings were obtained. For each of the 87 students, comprising the sample, a median achievement rating was determined. The higher the rating, the greater the degree of achievement or, conversely, the lower this rating, the greater is the degree of underachievement.

Mean achievement rating scores were computed for students whose responses placed them in one of the seven motive categories. Then, since there were few cases in certain categories, these were combined thusly: "O-E" and "O-mixed," "A-I" and "A-mixed," and "C-S" and "C-mixed." The significance of the differences between means of

TABLE I

REASONS COMPRISING THE L-M SCALE OF MOTIVES FOR GOING TO COLLEGE

Type of Motive	No. *	Reason
Occupational-Economic	1.	To get a better paying job
	4.	To meet the requirements of my professional choice
	7.	To help me choose a vocation
	10.	To keep from going to full-time work
	13.	To keep out of the draft
	16.	To have more influence in the community
	19.	To have a high standing in the family
	21.	To meet the requirements of a graduate school
	25.	To make contacts with people who might help me in my professional advancement
Academic-Idealistic	2.	To help satisfy my curiosity in an area of study
	5.	To satisfy my desire to learn
	8.	To be a more well-rounded person
	15.	To be of better service to society
	17.	To help me think logically
	20.	To help me develop a set of values
	27.	To be exposed to stimulating ideas
Conformity-Social	3.	To make friends
	6.	To participate in college social and athletic activities
	9.	To develop me socially
	11.	To enable me to get along better with others
	12.	To meet a future mate
	14.	To be like my friends since most of them are going to college
	18.	To please my parents
	22.	To please my teacher
	23.	To help me better understand myself and others
	24.	To follow the path of people whom I respect
	27.	To increase my status, because people nowadays are expected to get an education

*Order in which the reason was listed in the scale

TABLE II

DESCRIPTION OF SEVEN MOTIVE CATEGORIES

Designation of Motive Category	Description
O-E	Three "occupational-economic" reasons
A-I	Three "academic-idealistic" reasons
C-S	Three "conformity-social" reasons
O-mixed	Two "occupational-economic" reasons and one in any other category
A-mixed	Two "academic-idealistic" reasons and one in any other category
C-mixed	Two "conformity-social" reasons and one in any other category
O-A-C	One reason each from the categories "O-E," "A-I," and "C-S"

TABLE III

MEAN ACHIEVEMENT RATING SCORES RELATED TO REASONS REGARDED AS MOST INFLUENTIAL FOR GOING TO COLLEGE

Motive Category*	Men (N=38)	Women (N=49)	Total (N=87)
O-E	3. 01** (N= 3)	....	3. 01 (N= 3)
A-I	3. 38 (N= 3)	2. 99 (N= 7)	3. 10 (N=10)
O-mixed	3. 07 (N=17)	3. 03 (N=10)	3. 05 (N=27)
A-mixed	3. 25 (N= 7)	3. 25 (N=19)	3. 25 (N=26)
C-mixed	3. 13 (N= 3)	3. 00 (N= 1)	3. 09 (N= 4)
O-A-C	3. 12 (N= 5)	3. 19 (N=12)	3. 17 (N=17)

* Motive categories are described in Table II.
**The higher the score, the greater the degree of achievement.

TABLE IV

MEAN ACHIEVEMENT RATING SCORES RELATED TO REASONS REGARDED AS LEAST INFLUENTIAL FOR GOING TO COLLEGE

Motive Category	Men (N=38)	Women (N=49)	Total (N=87)
O-E	2.71 (N= 2)	2.21 (N= 2)	2.46 (N= 4)
C-S	3.00 (N= 4)	3.18 (N= 2)	3.06 (N= 6)
O-mixed	3.19 (N=14)	3.09 (N=20)	3.13 (N=34)
C-mixed	3.15 (N=15)	3.32 (N=23)	3.25 (N=38)
O-A-C	3.21 (N= 3)	2.42 (N= 2)	2.90 (N= 5)

TABLE V

t AND p VALUES OBTAINED IN TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEAN ACHIEVEMENT RATING SCORES OF MALE STUDENTS WHO REGARDED CERTAIN REASONS AS MOST INFLUENTIAL

Motive Category	N	Mean	Variance	t	p
Occupational-Economic	20	3.06	.19	1.6	> .05
Academic-Idealistic	10	3.29	.09		

TABLE VI

t AND p VALUES OBTAINED IN TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEAN ACHIEVEMENT RATING SCORES OF FEMALE STUDENTS WHO REGARDED CERTAIN REASONS AS MOST INFLUENTIAL

Motive Category	N	Mean	Variance	t	p
Occupational-Economic	10	3.03	.20	.79	> .05
Academic-Idealistic	26	3.18	.36		

TABLE VII

t AND p VALUES OBTAINED IN TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEAN ACHIEVEMENT RATINGS OF MALE STUDENTS WHO REGARDED CERTAIN REASONS AS LEAST INFLUENTIAL

Motive Category	N	Mean	Variance	t	p
Occupational-Materialistic	16	3.13	.138	.03	> .05
Conformity-Social	19	3.12	.150		

TABLE VIII

t AND p VALUES OBTAINED IN TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEAN ACHIEVEMENT RATINGS OF FEMALE STUDENTS WHO REGARDED CERTAIN REASONS AS LEAST INFLUENTIAL

Motive Category	N	Mean	Variance	t	p
Occupational-Materialistic	22	3.06	.345	1.6	> .05
Conformity-Social	25	3.31	.241		

combined motive categories was ascertained by applying t tests. The .05 level was adopted as the criterion of significance.

Findings

Tables III and IV present mean achievement rating scores of male and female college freshmen who regarded certain types of motives as most and least influential in their decisions to attend college. Data included in Tables V through VIII show the results of the tests of the significance of the differences between mean achievement rating scores. The results indicate that, for this sample, there exists no significant relationship between type of motivation for going to college and academic achievement.

Undoubtedly there are educators and college admission officials who think, a priori, that certain reasons for wanting to attend college are more desirable than others. For instance, when interviewing college applicants there may be a tendency to be partial to students who manifest "academic-idealistic" rather than "occupational-economic" points of view. It is generally agreed, however, that one crucial, personal criterion of college success is that the student achieves satisfactorily. This study suggests that "occupationally"-oriented students are just as likely to achieve commensurate with their estimated intellectual capacity as are "academically"-oriented and "conformity"-oriented freshmen. It seems, therefore, that a major determinant of academic achievement is not the quality or type of motive for attending college, but the intensity of the student's commitment to pursue his education to whatever end appears relevant at the moment.

In view of the importance attached to student motivation, it would certainly be worthwhile to replicate a study of this nature in other institutions to determine the generality of the conclusion drawn.

Summary

A sample of 38 male and 49 female college freshmen was selected, respresentative of the 1959 freshman class of Fairleigh Dickinson University, Teaneck, New Jersey. Students' motives for going to college were compared with their academic achievement, judged by at least five of their instructors. No significant relationship was ascertained between quality and type of motivation and academic achievement as measured in this study.

REFERENCES

1. Dyer, H. S. "A Psychometrician Views Human Ability," Teachers College Record, LXI (1960), pp. 394-403.
2. Fishman, J. A., and Pasanella, Ann K. "College Admission-Selection Studies," Review of Educational Research, XXX (1960), pp. 298-310.
3. Georgiana, D. W. S. "The Relationship of Two Experimental Measures of Student Motivation to Academic Success in College," Dissertation Abstracts, XV (1955), pp. 637-38.
4. Goldsen, Rose K., et al. What College Students Think (Princeton, N. J.: D. Van Nostrand Co., 1960).
5. Gowan, J. C. "Factors of Achievement in High School and College," Journal of Counseling Psychology, VII (1960), pp. 91-5.
6. Iffert, R. E. Retention and Withdrawal of College Students (Washington, D. C.: U. S. Department of Health, Education and Welfare, 1958).
7. Keezer, D. M. (Ed.) Financing Higher Education: 1960-70 (New York: McGraw-Hill, 1959).
8. Lang, G., Sferra, A. G., and Seymour, Marjorie. "Psychological Needs of College Freshmen Related to Their Academic Achievement," Personnel and Guidance Journal (in press).
9. Malsey, F. G. Construction of a Scale of Motives for Going to College, unpublished Master's Thesis, Fairleigh Dickinson University, 1960.
10. Michael, W. B., Jones, R. A., and Trembly, W. A. "The Factored Dimensions of a Measure of Motivation for College Students," Educational and Psychological Measurement, XIX (1959), pp. 667-71.
11. National Education Association and American Association of School Administrators. Higher Education in a Decade of Decision (Washington, D. C.: NEA, Educational Policies Commission, 1957).
12. Shaw, M. C., and Brown, D. J. "Scholastic Underachievement of Bright College Students," Personnel and Guidance Journal, XXXVI (1957), pp. 195-99.
13. Shuman, R. B. "College Drop-outs," Journal of Educational Sociology, XXIX (1956), pp. 347-350.
14. Slater, J. M. "Perception: A Context for the Consideration of Persistence and Attrition Among College Men," Personnel and Guidance Journal, XXXV (1957), pp. 435-40.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

EXPERIMENT IN INDEPENDENT STUDY (1956-1960)*

SAMUEL BASKIN
Antioch College

ANTIOCH COLLEGE has long sought to discover ways by which students could take a greater share of the responsibility for their own education.[1] It has tried to achieve this objective in a variety of ways: in its programs of work experience and education abroad, where students are almost completely on their own;[2] in its plan for community government, where students lead in planning and conducting many activities of the college program; in using students in freshman orientation and the advising program, in which upper-class students are the only adult residential advisers in freshman dormitories; and in the Administrative Council, where students take part with faculty members in making decisions on college policies and program.

While the College has sought to provide comparable kinds of responsibility-taking experiences within its academic program (through tutorial studies and credit by examination), these experiences have been largely limited to its more able students. For the most part, students have been expected to attend regularly scheduled class meetings, three or five times a week for 12 or 24 weeks, depending on the number of credits offered. Implicit in this scheduling was the assumption that students needed at least this many meeting hours with the instructor to learn effectively. But did they?

Beginning in 1956, the College began to examine this assumption more closely. It proposed controlled group studies to test various methods of independent study in regular teaching. The purpose was to see whether these independent study methods could help students take more responsibility for their own learning--and whether the College might not only increase the quality of students' education, but also save instructional time.

Hypothesis

Two principal hypotheses were proposed:

(1) That students in the independent study courses would learn as much as those taking courses using regular lecture-discussion instruction.

(2) That students in the independent study courses would be as satisfied with the courses as those taking courses using the regular lecture-discussion methods.

The College also sought to study the amount of instructional time used under both methods of instruction.

Plan of the Studies

The experiment called for setting up control and experimental groups for most of the courses in this study.[3] Both groups were to be taught by the regular and the experimental methods. Both groups were to be taught by the same instructor, and both were matched on a number of variables, including field, year in college, sex, and background knowledge at the time of taking the courses. Students taking the courses were expected to cover the same course materials and to achieve the same course objectives as outlined in previously prepared syllabi. The principal difference between the groups was that the experimental groups used a period of independent work. This period varied from course to course; the independent study groups averaged 20 to 30, and in some instances 40, fewer class hours than did those taught the regular way.[4]

The Instruments of the Study

Pre-tests used to determine the comparability of the groups included the Yale I Educational Aptitude Test, as a measure of the students' scholastic ability; the Yale 5 Educational Aptitude Test, as a measure of mathematical ability; and several course examinations, including both subject-matter and task-oriented tests, as a measure of students' course knowledge at the time of beginning the course. In addition, members of the control and experimental groups were studied for their comparability on each of the following variables: field, sex, year in college, and rank in high school.

Post-course measures included multiple-choice and short-answer tests, to determine content learning, and essay, task, and laboratory problem assignments, to measure critical thinking ability and ability to handle complex problems.

Two measures were used to obtain evidence of the students' satisfactions and dissatisfactions with the teaching procedures: a teacher rating scale (Survey of Student Opinion of Teachers) which has

* All footnotes will be found at end of article.

been used at Antioch since 1947; and a student satisfaction scale developed for the study to get information about students' reactions to the teaching procedures used.[5]

In addition, instructors were asked to keep records of the time used in teaching the different ways. These records, filled out weekly, reported time spent in class, in conferences with students, in preparation for classes, and in evaluation of student work. The smallest unit of time used was a quarter of an hour.

The Independent Study Courses

Organization

Instructors were given considerable latitude in the method of independent study used. Some instructors used principally teams or small groups of five to ten students, who did their autonomous work together. Some instructors used individual study approaches, or what came to be known as the "lone wolf" method of independent study. Still others used combinations of small group and individual independent study.

Instructor-Student Meeting Patterns

Instructors were free to choose how they wished to organize their courses. Sometimes they asked students to do some independent work each week and meet once or twice a week with the instructor (in contrast to the control group, which met with the instructor for all regularly scheduled classes). Other instructors used "block units," meeting the experimental groups regularly during the early and latter parts of the course and scheduling blocks of completely independent work for one, two, or three weeks at intervals in the course. Still other instructors, depending on the instructional material and student needs, used different combinations of regular class meetings and independent work.

Training Procedures

Considerable attention was given to preparing students for working on their own. Training materials developed for this purpose included discussion outlines and thought questions for the autonomous study groups, reference materials, and specially prepared instructional materials. Among the latter were aids designed for chairmen of small groups on how to lead a discussion, suggestions for discussion recorders, and suggested study tactics for individuals working on their own or in small groups.[6] In addition, a number of instructors conducted demonstration sessions on effective and ineffective work in small groups and held post-class conferences with the chairmen and recorders of small groups. While instructors spent different amounts of time in training groups or individuals for working on their own, some kind of training aid was prepared for each independent study group.

Results

Achievements of the Group

(1) Students learned equally well whether they had studied under the regular method of instruction or under the experimental method of independent study. No significant pattern emerged favoring one group over the other. This held true whether multiple-choice examinations, short-answer tests, essays, or task assignments were used to evaluate the results. It held true for each year of the study, and it held true for both beginning and advanced courses.

(2) Data on retention of learning (two years after completion of the course) revealed no significant differences favoring one group over the other.

(3) There was no evidence that the independent study methods needed to be reserved for the superior or advanced student only. Students at differing levels of ability and at different year levels did well (or poorly) with these teaching methods. Academic potential, at least within the ranges studied at Antioch,[7] did not appear to be the sole determiner in the success or failure of these methods, and there were some indications that personality and attitudinal factors might be as important in learning by these methods as individual ability.

Satisfaction with Control and Experimental Methods

While students still tended to prefer the lecture-discussion teaching, they expressed a growing acceptance of and satisfaction with independent study as they became more familiar with it. The study data showed a decided shift in student attitudes from one of general dissatisfaction with independent study in 1956, to few differences in the degree of satisfaction under either method of instruction in 1960. There is little question, as analyses of the student satisfaction scales and conferences with students indicated, that students come to college largely unprepared for independent learning. If independent study methods are to succeed, students need a great deal of help in learning how to work on their own.

Use of Instructional Time

The study data revealed that there were few savings in instructional time during the early years of the study because of the additional time needed to prepare lesson material for the independent study groups. There was some indication, however, that time might be saved in some courses once the instructor had mastered the use of inde-

pendent study and had completed the cycle of preparing special materials. This was especially true for several courses in the last two years of the experiment, including two philosophy courses in which the instructor saved up to 57 percent of the total time that normal procedures would have required for the same number of students.

Conclusions

Although an instructor satisfaction scale was not used, instructor reactions to the experimental procedures were secured through written reports and individual conferences. Perhaps the best evidence of their feelings is the fact that most instructors (especially those participating in the studies since 1957) have continued to use independent study procedures, and in the fact that an increasing number of faculty members, including many who did not participate in the experimental studies themselves, are now offering students the opportunity for on- as well as off-campus independent study.

Antioch College does not view these new methods of instruction as a panacea for higher education's twin problems of quality and quantity. Nor does it view them as a glorified do-it-yourself plan that works simply by turning the student loose on his own; for the instructor must carefully select and structure the student's learning experiences. The instructor's job, although it may be different, is no less difficult.

While it is not clear that the experimental methods will result in a net saving in instructional time, for those who have been engaged in the study this fact in and of itself is not the significant point. The real promise of the studies is in the shift they may bring with them in the educational orientation of both students and teachers. Most importantly, they argue for the far greater use of the student's own resources for learning. They press us to shake ourselves loose from some of our educational habits, and they offer new assurances of the unused potential for learning—and one that we have hardly begun to tap-- that exists within the student himself.

FOOTNOTES

1. This report was originally presented as Antioch College Reports, No. 2, March, 1961. The studies reported herein were supported by grants from the Fund for the Advancement of Education.
2. Under Antioch's educational plan of study-plus-work, students alternate their time between on- and off-campus experiences. As part of the off-campus experience, students hold regular jobs at regular pay. The program presently includes over 450 co-operating employers in 30 states. The College has recently extended this concept of education still further so that students are able to spend a full year abroad as part of their degree program. About 150 students are expected to take part in this program in some 20 countries this year.
3. Of 19 courses studied from 1956 to 1960, 13 were taught by both experimental and control procedures.
4. The reduction in instructional time ranged from 34 percent to 61 percent, with a median of 46 percent.
5. In addition to these measures, two other tests of skills in and attitudes toward independent study were used in the 1957-58 study. These were a Situational Problems Test developed at Antioch and five scales of the California Personality Inventory. These measures are described in more detail in the 1957-58 study report. See also Ruth Churchill, Evaluation of Independent Study in College Courses (unpublished doctoral thesis), 1960, University of Minnesota.
6. A number of these aids are being brought together in a manual on independent study that the College is preparing (Morris Keeton and Samuel Baskin, Independent Study: A Teacher's Guide) to help other faculty members train students for working on their own.
7. These results were not confined to Antioch alone. Surveying the results of some sixteen institutions (both public and private) that had engaged in similar research, a report of the Fund for the Advancement of Education notes that "Almost without exception, the customary academic examinations showed that students in the independent study experiments learned at least as much as the students who had regular class work. Rarely were there statistically significant differences in the performance of the experimental and the control groups on regular or special examination."

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

A COMPARISON OF KNOWLEDGE OF HIGH SCHOOL SUBJECTS POSSESSED BY COLLEGE APPLICANTS AND NON-APPLICANTS

EUGENE S. EDGINGTON
Kansas State Teachers College, Emporia

THIS STUDY compared Kansas State Teachers College applicants with non-applicants in regard to knowledge of high school subjects. The two groups were equated for ability.

Measuring Instruments Used

The measure of ability used in this study was the score on the General Ability test, a component of the Kansas High School Senior Comprehensive Examination battery, published by the Bureau of Educational Measurements, Kansas State Teachers College. The measure of knowledge of high school subjects was the Total Achievement score, which is a composite derived from scores on the subtests of the Kansas High School Senior Comprehensive Examinations: Science, Mathematics, English, and Social Studies.

Subjects and Method

The applicants were the 707 first-semester freshmen coming to Kansas State Teachers College in 1960 from Kansas high schools at which they previously had taken the Senior Comprehensive Examinations. The Total Achievement score of each of the 707 freshmen was compared with the median Total Achievement score of those high school seniors who took the Senior Comprehensive Examinations at the same time and obtained exactly the same General Ability score as the freshmen. The median Total Achievement score was not the median for non-applicants to Kansas State Teachers College, but was the combined median for applicants and non-applicants. If students with the same General Ability scores as those of the 707 entering freshmen were randomly selected from the 17,007 high school seniors who took the Senior Comprehensive Examinations when the entering freshmen did, about half of the students would be expected to be above the median Total Achievement score and half below.

To compute the medians, the Total Achievement scores were grouped into intervals with an interval length of one score unit. Thus, in effect, scores were spread uniformly throughout the interval containing them, even though all the obtained scores within an interval were identical in size. Each Total Achievement score falling in the interval containing the median was considered to be partly below the median and partly above the median; the fraction of the score considered to be above or below the median corresponded to the fraction of the interval falling on that side of the median.

Results of First Analysis

The table at the top of the next page shows, for each General Ability high school decile, the number of freshmen above and below the high school medians. The total number of freshmen scoring below the median Total Achievement score for high school seniors possessing this particular General Ability score was 305.58 and the number above the median, 401.42. Using an expected value of $N/2 = 353.5$ for the number above or below the median, and making a correction for continuity, the binomial test yielded a $\underline{z}$ of 3.6. Normal probability curve tables show that the one-tailed probability of getting such a large $\underline{z}$ by chance is .0002. In other words, only two times in 10,000 would such a preponderance of students scoring above the median occur from sampling error. The conclusion is that the superior knowledge of high school subjects possessed by Kansas State Teachers College freshmen compared with high school graduates of the same ability is not to be attributed to chance, but instead indicates the existence of special determining factors. This is especially interesting in view of the fact that Kansas State Teachers College is required by state law to accept any applicants who are Kansas high school graduates; the college admission policy cannot account for these results.

NUMBER OF COLLEGE FRESHMEN ABOVE AND BELOW MEDIAN TOTAL ACHIEVEMENT SCORES FOR HIGH SCHOOL STUDENTS OF EQUAL ABILITY

High School General Ability Decile	Number of Freshmen	Number Below Median Total Achievement Score	Number Above Median Total Achievement Score
1	18	6	12
2	14	2	12
3	40	14	26
4	59	22	37
5	52	19	33
6	61	23.23	37.77
7	89	41.88	47.12
8	112	47.71	64.29
9	145	70.08	74.92
10	117	59.68	57.32
Total	707	305.58	401.42

Results of Second Analysis

A comparison of the number of freshmen above and below the median was made for the freshmen from the lower seven high school General Ability deciles. Of those 333 freshmen, 128.11 were below the median and 204.89 were above the median. This provides a z of 4.1, for which the one-tailed probability of chance occurrence is .00003. On the other hand, of the 374 students in the 8th, 9th, and 10th deciles, 177.47 were below the median, and 196.53 above, yielding a z of .94, which is so low as to occur 17 percent of the time by chance, so the difference in number above and below the median is not significant even at the .10 level.

The results of this study are consistent with the assumption that ability is not a factor in determining whether a Kansas high school graduate will become a Kansas State Teachers College applicant, except indirectly through its effect on his knowledge of high school subjects. The table shows that more people from the higher levels of General Ability entered Kansas State Teachers College than from the lower levels. This can be explained in this way: more people in the upper levels of General Ability have good Total Achievement scores than in the lower levels. In the upper levels of General Ability, the median Total Achievement scores are so high that people below the median, as well as those above, generally have good Total Achievement. But the median Total Achievement scores for the lower levels of General Ability are so low that a good Total Achievement score almost certainly would be above the median.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

COLLEGE PREPARATORY COURSE WORK*

GERALD H. WHITLOCK
JAMES R. MONTGOMERY
The University of Tennessee

FOR SEVERAL years The University of Tennessee has offered special summer courses for high school graduates planning to become college freshmen.[1] A student may elect to take any one (but not more than three) of the following non-credit courses (25 hours of classroom time each): "How to Study," "Developmental Reading," "Preparatory English," or "Preparatory Mathematics."

What effect, if any, do these courses have on college-bound students? Do they, as might be assumed, assist a pre-freshman in his later college work?

Since individuals enrolling do so in order to help themselves in their freshman year, it seems reasonable to expect that such students would profit by the courses. At least they should do no worse in college than their counterparts, the freshmen who do not take these courses.

Freshman orientation courses which give information on study habits, reading practices, and use of a library are found in many institutions.[2] Based on the findings at The University of Tennessee, it seems that a note of caution is advisable and that such work should be evaluated in order to assure that the course is really helping the students involved.

Problem

This study sought to determine the effect of the summer program on first quarter college grades of the students involved. Those enrolled in the summers of 1959, 1960, and 1961 were included in the review. Students who took one or more of the pre-freshman courses and later entered The University of Tennessee (the Experimental Group) were compared with a matched group who did not take the courses (the Control Group).

For each individual in the Experimental Group a person was selected as a member of the Control Group who had the same ACT score,[3] same sex, entered the same college, and either lived on or off the campus. Efforts were made to keep the high school grade-point averages close, but, as will be seen, in certain cases a significant difference occurred in them.[4]

Comparing the grades of the Experimental and Control Groups at the end of the first quarter in The University of Tennessee as full-time freshmen, the difference in achievement could be obtained. Using a "t" test, with the matched pair design, the significance of the difference in grade-point average was determined for the individual course (in the case of "Preparatory English" and "Preparatory Mathematics") or overall college grade-point average (for "How to Study" and "Developmental Reading" courses).

Results

In Table I is presented a review of the grade-point averages of the Experimental Group and the Control Group. The Experimental Group took the course in improving study methods. Although the Control Group in 1959 had a higher high school grade-point average ($<.05$), an analysis of covariance was not necessary since the correlation between high school grade-point average and first quarter overall grades was not significant. Data in Table I indicate that students taking the "How to Study" course consistently scored lower in their first quarter of college academic work than did those in the Control Group. Findings in the University of Tennessee survey do not agree with the research reported on an apparently similar course conducted in another state university where there was an immediate and continuing increase in scholastic gain for students taking a study skills course.[5]

The results of a course in "Developmental Reading" on the Experimental Group as compared with the Control Group without the course are reflected in Table II. After 1959, in which there appears an unusually high grade-point average in the Control Group at least in relation to other college grades shown in this study, no significant difference appears between the two groups although the Control Group's mean grade-point average consistently runs higher. While previous studies have brought out the fact that remedial reading programs can improve scholastic standing, not particularly specific course grades, the fact that such studies have been

* Footnotes will be found at end of article.

TABLE I

COMPARISON OF THE EXPERIMENTAL GROUP IN THE "HOW TO STUDY" COURSE WITH THE CONTROL GROUP ON THE BASIS OF FIRST QUARTER OVERALL GRADE-POINT AVERAGE

Year	N	Control Mean College GPA	Experimental Mean College GPA	t	Significance
1959	9	1.95*	1.22	2.18	< .05
1960	15	1.84	1.43	2.58	< .025
1961	16	2.04	1.65	4.51	< .01
Combined	40	1.95	1.47	4.07	< .005

*All grade-point averages expressed in terms of a four point system where "A" = 4 and "D" = 1.

TABLE II

COMPARISON OF THE EXPERIMENTAL GROUP IN THE "DEVELOPMENTAL READING" COURSE WITH THE CONTROL GROUP ON THE BASIS OF FIRST QUARTER OVERALL GRADE-POINT AVERAGE

Year	N	Control Mean College GPA	Experimental Mean College GPA	t	Significance
1959	6	2.62	1.89	3.42	< .01
1960	10	1.61	1.37*	1.23	N/S*
1961	9	1.92	1.63	1.31	N/S
Combined	25	1.96	1.59	2.94	< .01

*Although the Experimental Group had a significantly higher high school grade-point average (< .01), an analysis of covariance was not necessary, since the correlation between high school grade-point average and first quarter overall grade-point average was not significant.

TABLE III

COMPARISON OF THE EXPERIMENTAL GROUP IN THE "PREPARATORY ENGLISH" COURSE WITH THE CONTROL GROUP ON THE BASIS OF FIRST QUARTER GRADES IN FRESHMAN ENGLISH

Year	N	Control Mean College GPA	Experimental Mean College GPA	t	Significance
1959	10	1.60*	1.20	1.08	N/S*
1960	13	1.54	1.77	-1.14	N/S
1961	16	1.25	1.13	.50	N/S
Combined	39	1.44	1.36	.50	N/S

*Although the Control Group had a higher high school grade-point average (< .05), an analysis of covariance was not necessary, since the correlation between high school grade-point average and first quarter English grades was not significant.

TABLE IV

COMPARISON OF THE EXPERIMENTAL GROUP IN THE "PREPARATORY MATHEMATICS" COURSE WITH THE CONTROL GROUP ON THE BASIS OF FIRST QUARTER GRADES IN FRESHMAN MATHEMATICS

Year	N	Control Mean College GPA	Experimental Mean College GPA	t	Significance
1959	4	1.75	.50	2.61	< .05
1960	4	2.25	.50	3.66	< .025
1961	7	1.86	1.86	0	N/S
Combined	15	1.93	1.13	2.18	< .025

evaluated over longer periods of time may influence the outcome.[6] The students in The University of Tennessee who took "Developmental Reading" did not show evidence of an immediate advantage in scholastic success over their counterparts; in fact, they showed a slight negative reaction.

Results from "Preparatory English," as seen in Table III, indicate no significant differences among the groups. Those who took the course in "Preparatory English" performed in college English in almost the same manner as the Control Group. In 1960 the Experimental Group did slightly but not significantly better. Since the course closely parallels the content of freshman English, lack of success is somewhat surprising. Students taking it, however, are not receiving any detrimental effects.

Although the numbers involved are small, students enrolled in the "Preparatory Mathematics" course (see Table IV) performed two years out of three significantly lower than the Control Group in freshman mathematics. It appears safe to say, therefore, that pre-freshmen electing this mathematics course will do no better than students who do not take it.[7]

Summary

At first glance the results of these pre-freshman courses at the University seem to leave much to be desired; however, the indications are that there is a much greater and more fundamental problem involved. Students elect to take these courses, not knowing whether they are really deficient in the skill involved or not. Having taken the skill course, they either fail to improve or perform worse than a control individual.

Possibly a selecting process is occurring which brings a certain type of student to take these courses, thereby establishing a variable which this research failed to control. These students, for example, give up various other summer plans in order to enroll in the courses which they elected to take, which their parents believed they needed, or which some other person had advised them to take. Therefore, either the student or someone in whom the student has confidence, has indicated that they believe the prospective freshman is weak in a given subject or at least has room for improvement in certain skills before entering college. If this theory is right, students in the Experimental Group may have a feeling of inadequacy or lack of self-confidence which is strengthened and confirmed by their initiation into college work. Another possible theory is that pre-freshmen enrolling expect their summer's course-work experience to serve as a panacea for all their academic difficulties. (A study at The University of Tennessee is being designed to test these theories.)

Students who take preparatory courses and later enroll in The University of Tennessee apparently receive little immediate benefit from these courses as far as grades and grade-point averages are concerned. Of course, if in actuality such persons are influenced in their college success by non-intellectual factors, such as self-confidence, then there may have been improvement by the Experimental Group which this research did not record.

This study suggests that the teaching-learning situation in elective preparatory courses may need to be reviewed to see if there are certain other approaches in teaching these classes which might produce results different from those obtained in this study. Different teaching-learning situations might be more suitable for the type of individuals enrolling.

FOOTNOTES

* Some of material in this paper was presented to the Nashville Conference on Research related to college admissions sponsored by the Southern Regional Education Board, August 1962.

1 College credit is not given for these courses. The Division of University Extension, which administers these courses, uses regular University faculty members to teach the courses.

2 George H. Greene, "Freshman Orientation Courses in Small Colleges," Personnel and Guidance Journal, XXXII, No. 2 (April, 1954), pp. 480-482.

3 Other research in The University of Tennessee has shown that high school grades or American College Testing scores, when used individually, predict college grades equally well (only one point difference in favor of high school grades). See E. E. Cureton, "Background Characteristics of the Freshman Class Entering The University of Tennessee in September, 1960," Mimeographed, 1962.

4 The mean high school grade-point average for the Experimental Group was 2.5, which is significantly lower than the total freshman mean grade-point average of 3.0 (in the fall of 1961), but not significantly different from that of the Control Group (also 2.5).

5 James G. Shaw, "An Evaluation of a Study Skills Course," Personnel and Guidance Journal, XXXIII (April, 1955), pp. 465-468.

6 John R. Wittenborn, "Classes in Remedial Reading and Study Habits," Journal of Educational Research, XXXVII, No. 8 (April, 1944), pp. 571-586.

7 In a somewhat similar study made at Cornell University some years ago, no significant difference appeared between a control and experimental group, although those involved believed the course definitely served to help the students enrolled. J. S. Ahmann and M. D. Glock, "An Evaluation of the Effectiveness of a Freshman Mathematics Course," Journal of Educational Psychology, L (February 1959), pp. 41-45.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

STUDY OF UNDERGRADUATES OF UNIFORMLY HIGH ABILITY

PAUL S. BURNHAM and ROBERT R. RAMSEY, Jr.
Yale University

OPPORTUNITIES for increasing selectivity in undergraduate admissions plus a growing number of national programs for the identification of the talented have resulted in the presence in recent Yale classes of a number of very able students with relatively equal promise for work in the Humanities and Sciences. Interest in the undergraduate careers of these promising young scholars, and in particular the curricular programs they chose, led the Office of Educational Research at Yale to study a sample of uniformly able students from the 1959 and 1960 classes; its findings are summarized in this report.[1*]

Criteria for Sample Selection

The sample consisted of 48 men (Yale College and School of Engineering combined); 23 from the class of 1959, of whom 18 graduated in 1959, and 25 seniors from the class of 1960. At matriculation, each student had a prediction of 80 or higher on a scale on which the average for all matriculants was approximately 77. In addition, each student had high College Board scores on two verbal and two quantitative entrance examinations.[2] There were more independent than public school graduates (27 as compared to 21), and roughly two-thirds were recipients of financial aid from University or outside scholarship programs.

The 1959 sample provided statistical data on most recent graduates while the 1960 sample furnished both statistical and interview data on seniors. Since Yale's permissive entrance examination requirements limit the possibilities for indentification of student abilities in the quantitative area, the 1959 sample had to be drawn from only 50 percent of the class since only 501 of the 1,002 matriculants had offered the four-year mathematics achievement score for entrance. Similarly, the 1960 sample had to be drawn from the 55 percent who had offered mathematics achievement scores at entrance.

Academic and Extra-Curricular Performance

Of those completing three years' work, two-thirds did so at a level of 85 or above while roughly only one-sixth produced non-honors work. Three-year averages for 60 percent of the sample placed them in the top 10 percent of their class. Of the 18 members of the class of 1959 graduating on schedule, 12 earned graduation honors, four were elected to Phi Beta Kappa, and four to Sigma Xi.[3]

Extra-curricular interests of the group were heavily weighted in favor of the non-athletic. There was some participation in intramural sports but none in major varsity sports. Debating, political organizations, professional societies, literary and journalistic groups, and service organizations were the most popular extra-curricular activities.

Fields of Undergraduate Major

Before coming to Yale, two-thirds of the sample expressed a desire to major in Engineering or one of the Sciences (Table I). Pre-matriculation percentages for the three broad divisions were as follows: Science and Engineering, 69 percent; Arts and Humanities, and Social Sciences, 25 percent (6 percent were uncertain about their probable college major). In college, however, the sample exhibited a marked shift away from the Sciences and Engineering. Final percentages were: Arts and Humanities, and Social Sciences, 63 percent; Sciences and Engineering, 37 percent. The 1959 sample contained one Scholar of the House (Physics). There were no Scholars of the House in the 1960 sample.

In an attempt to discover "why" these students of uniformly high abilities followed one curricular direction rather than another, the 25 seniors were interviewed. Designed as an informal, open-ended discussion of their formal intellectual experience, the interviews focused on the origins and development of current educational and vocational plans.

---

* Footnotes will be found at the end of the article.

TABLE I

CHOICE OF MAJOR

Major	Pre-Matriculation Choice		Final Choice	
Arts and Humanities	7	(3)	20	(12)
History and Social Sciences	5	(3)	10	(4)
Sciences and Engineering	33	(18)	18	(9)
Uncertain	3	(1)		
Total Study Sample	48	(25)*	48	(25)*

* Numbers of Seniors interviewed (Class of 1960) are shown in parentheses.

TABLE II

CAREER GOALS BEFORE AND DURING COLLEGE

Field	Career Choice Before College	Career Choice Now: Humanities and Social Science Majors	Career Choice Now: Science and Engineering Majors
Law	3	7	-
Medicine	4	3	1
Science or Engineering	12	-	5
Business	1	-	-
College Teaching	1	2	-
Ministry	1	1	-
Uncertain	3	3	3
Total	25	16	9

TABLE III

ATTRACTIVENESS OF MAJOR*

Reasons Given	Sixteen Humanities and Social Science Majors	Nine Science and Engineering Majors
Broad enough to allow for other electives	4	-
Developing earlier interests	2	5
Good teaching	2	1
Embraced concerns of all others (History)	2	-
Disillusionment with other fields	1	3
Possibility for original contribution	-	1

* Number of responses rather than number of students.

These semi-structured sessions, varying in length from 20 minutes to an hour, sought information relative to questions such as the following:

(1) What were your vocational goals before college?
(2) Were these reinforced or discouraged by your parents? By your teachers? By other acquaintances?
(3) At college were you attracted to (undergraduate major) because of subject content, quality of instruction, or other factors?
(4) What courses of study have you liked least? Why?
(5) Has semester or summer employment helped to crystallize your career plans?
(6) Has counseling at college been a significant influence in your choices?
(7) How do you expect to use your undergraduate training?
(8) In broad overview what person(s) has had the greatest impact on your career thinking?
(9) If you were beginning again, would you do things differently?
(10) Do you think you would have chosen differently at another college?
(11) Are you generally satisfied with your educational and career choices to date?

Since most subjects in this sample seemed to think in terms of a Humanities-Science dichotomy, their comments have been grouped accordingly. It will be apparent, however, that small numbers preclude sweeping generalizations. Examination of certain responses (Table II) reveals a variety of career goals, both before and during college. As has been observed frequently, there was a noticeable shift away from Science and Engineering and, in this sample, a trend toward Law or slightly greater career uncertainty.

Parents, high school teachers, and counselors continued to be the main influences on the student's pre-college educational and career plans. Few reported experiencing a direct push from their parents toward a specific educational or vocational goal, but nearly all recognized some kind of subtle or covert influence that "probably rubbed off indirectly."

Table III suggests that choice of major for this sample was more a function of subject content than methodology or techniques. Humanities and Social Science majors took comfort in the opportunities for broad, free-ranging intellectual adventure while Science majors were most often following through on interests stimulated initially in earlier work in school. For many, the attractiveness of the Humanities and Social Sciences was the practical consideration that these fields were "broad enough to keep the largest number of avenues open."

Semester or summer work experience had not been a significant factor in shaping the educational choices of the sample, though a few mentioned its value as a "reality testing" device. Seeing what the actual outside world of work is like caused a few shifts of emphases within fields (e.g., from applied to theoretical interests), but almost no changes between fields have occurred as a result of these experiences.

Less than one-fourth of the sample found counseling or advising an important factor in their curricular and career choices, which reinforces other impressions that this highly able group is typically self-sufficient when educational decisions are to be made. In talking with these men, one comes to realize that they are not only highly able but also are equally stable and independent in their overall orientation to daily life. Time after time there was evidence that their choices were products of personal thought and introspection rather than the issue of a single influence or event.

When asked if they would do things differently, given another chance, nearly all indicated they would probably choose the same department or area again. Representative of their general feeling is the following: "I would still stay in . . . , though probably change a few of my courses in that department. Also, I would try to take a few more courses outside the department. It bothers me a little that it's a pretty secular program."

A large proportion of the 25 interviewed expressed regret at the "heavy specialization requirements" in most majors and further lamented that "there is no premium on the versatility of knowledge." A few expressed interest in broader, if not deeper, knowledge, and one felt a four-year Directed Studies Program might be a useful addition to Yale offerings.

Scientists in the sample were frequently critical of laboratory report writing, labeling it onerous, overdone, and just plain busy work. Some reported that it had had a negative, stultifying impact on their intellectual interests and outlook.

Members of the sample felt they would have majored in the same or closely related subjects at another institution, and, the above exceptions notwithstanding, nearly all expressed general satisfaction with their overall college experience.[4] A sampling of their spontaneous comments provides a final touch of flavor:

> Freshmen need more of the better teachers. As things now stand, you finally get a big man at the end of Junior year; and just when your intellectual excitement gets underway, graduation arrives. (Humanities Major)

There is too little concern for the "teaching" role. Some junior men would make excellent teachers if they were not under the terrific scholarship role, producing a book between two covers that ends up on somebody's shelf and is never read. (Humanities Major)

The atmosphere is definitely toward the liberal arts. There's a hierarchy of English honors majors at top, then History honors majors, English standard majors, and so on. In the dining hall, for example, you sense this as you sit at the table. (Mathematics Major)

Science majors are at a disadvantage when it comes to classroom hours. My 26 hours per week in the lab can be contrasted with the 12 hours of my non-science friends. (Chemistry Major)

Concluding Comment

One is impressed with the variety of reasons given for their decisions by these very able students, by the shifting from earlier plans, and by the deterioration in number of prospective Science and Engineering majors; but fundamentally, the satisfactions expressed by these extremely able students are much more impressive than the shifts in plans or dissatisfactions. They were offered freedom of choice, and on the whole they found major programs which were intellectually satisfying.

In summary, these extremely able undergraduates appeared to be stable, self-sufficient individuals who chose their college majors independently of extraneous considerations and primarily on the grounds of likelihood for maximum intellectual fulfillment. By virtue of balance in the distribution of their high level talents, they were really free to choose any program of concentration. Colleges and universities should have more such students, and in turn they must recognize the obligation to provide commensurate challenges and opportunities in all areas for the uniformly able students, particularly those of high ability.

FOOTNOTES

1. Earlier studies on related topics were the following two:
   A. B. Crawford, "Survey of Students' Elections in Science," 1939 (Mimeo).
   A. B. Crawford and Paul S. Burnham, "Science Students in the Class of 1957," Research Division, Office of Counseling, Placement and Research, Yale University, 319 pp., 1958 (Mimeo).
2. Range of CEEB scores for members of the sample, 666-800; means on two verbal tests, 720, and on two quantitative tests, 728.
3. The five drop-outs (Class of 1959) were accounted for by accidental death, switch to a five-year engineering program, medical leave of absence, low standing, and resignation to go to work.
4. Similar satisfactions were reported by Class of 1957 seniors (Yale) interviewed in an earlier study (Crawford and Burnham, op. cit., pp. 25-26).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

COLLEGE SCHOLASTIC PROGRESS PATTERNS AS PREDICTIVE INDICES FOR COUNSELING

F. CHANDLER YOUNG
University of Wisconsin

THE PROBLEM of being misled by predictive information is especially important to American higher education at this critical period of rapid expansion. Prospective and currently enrolled college students need realistic counseling in regard to their college plans. This is a report of the development of some college scholastic progress patterns which may be more valid and more useful as predictive information for counseling than that which is presently available.

The basic idea underlying the development of this information is that of describing the actual in-college performance of former students of a college and then using this description as a pattern for the expected progress of prospective or currently enrolled students of that college. The description of former students was obtained by selecting various homogeneous groups of former students and then charting the progress of each group over a four semester period. The progress of each group included both an achievement aspect and a persistence aspect. Achievement is shown by the proportion of each group in each of four grade-point average ranges. The persistence aspect is shown by the proportion of each group leaving at various times over a two year period. Whether or not these departures are voluntary is also shown.

The development of these progress patterns began with two basic premises: 1) that the two factors most significantly related to college scholastic progress are previous academic achievement and scholastic ability; and 2) that the scholastic progress made by prospective or currently enrolled students of a college will be essentially the same as progress made by previously enrolled students of that college.

Many studies have reported the relationship of the two factors, ability and previous achievement, to college grade-point averages. Travers[1*] has reviewed predictive studies of this kind at all educational levels, and Garrett[2] has done so for predictions in liberal arts colleges.

An aspect of progress other than college grade-point average, however, is that of staying power or, stated negatively, tendency to leave college. Table I presents some evidence to show the significance of the relationship of the two factors, ability and previous achievement, to staying power or tendency to leave college. In Table I, consider the men of low scholastic promise, that is, those who scored among the lowest third on the College Qualification Test and who earned less than a C average the first semester: 29.2% of this group left voluntarily before the junior year as compared with 16.9% of the men with high scholastic promise; likewise 54.6% of the low scholastic promise group was dropped as compared with only 1.2% of the high scholastic promise group. The effect of previous achievement and scholastic ability on departures is also shown for the women. In the case of women, however, ability and previous achievement have little effect upon voluntary departures.

Evidence in support of the second premise, concerning the relationship of the progress of currently enrolled students to former students, is presented in Tables II and III. Table II shows the similarities from year to year for the grade-point average aspect of progress. There was little variation in mean grade-point averages from year to year over a 10 year period. For example, consider the second column in Table II. This column shows 10 different groups of freshman men in the College of Letters and Science, each group being present for the second semester in the year shown. The range of the mean grade-point averages shown in the third column is 2.22 to 2.32. There is no trend upward or downward in terms of grade-point average, but there is an upward trend in the size of the groups, reflecting enrollment increases. Each of these groups is, of course, quite heterogeneous in respect to academic background, scholastic ability, and many other factors. The columns for sophomores, juniors, and seniors similarly show little variation in mean grade-point averages from year to year. Furthermore, differences between the mean grade-point averages of the four classes presented are essentially the same from year to

* Footnotes are to be found at the end of this article.

year. In each year, for example, the sophomores earned higher grades than freshmen, juniors higher than sophomores, and seniors higher than juniors.

Table III shows the similarities in departures from year to year by comparing a 1958 and a 1959 entering group. The basic patterns of departure are essentially the same, and differences in kind and time of departures are slight. For example, in each year the greatest exodus was after the second semester, and the dropped and voluntary departures are about equal, there being a slightly higher proportion leaving voluntarily than being dropped.

Accepting the two basic premises previously discussed, scholastic progress patterns were developed from the records of former students of the College of Letters and Science at the University of Wisconsin for use with students currently enrolled in that college. Some 9,000 new undergraduate students entered the University of Wisconsin in September of 1958 and 1959. Data regarding test scores, credits carried, grade-points earned, nature of departures, and other information were obtained for these students. All freshman men with no previous college record who first entered the College of Letters and Science and completed the first semester were selected from these 9,000. Those excluded were all women and those men who had had some previous college work, who had entered some college or school of the University other than the College of Letters and Science, or who had registered for but had not completed the first semester. A very few were excluded because they had not taken the College Qualification Test. This selection procedure resulted in a total of 1,924 students.

The progress of this group of 1,924 men students is shown in Chart I. The group is homogenous in respect to sex, previous level of education, and general educational plan. However, it is quite heterogeneous in regard to high school rank, first semester grade-point average, scholastic ability, and, of course, individual characteristics.

Chart I shows the progress of this group through four consecutive semesters, each circle representing a semester. The achievement aspect of progress is shown by the proportion of the group achieving in each of four grade-point average ranges in each of four consecutive semesters. The solid areas show the honor range, the two dotted sections the average range, and the lined areas the unsatisfactory range. The clear area represents the proportion of the original group not completing that semester. The departure aspect of progress is shown by the figures below each circle. These figures are percentages of the original number of 1,924 men and show when the departures occurred and whether or not they were voluntary.

Each circle in Chart I is designed to show proportions of a group of students in respect to both the original entering group and to that part of the original group completing each semester. In order to illustrate this dual reference of each proportion, consider the circle in the lower right hand corner, representing the fourth semester. Visual inspection shows that a little less than two-thirds of the original group completed the fourth semester. The proportion achieving in the 3.00 to 4.00 GPA range, the solid section, can be seen both in relationship to the original group and to the group which completed four consecutive semesters. In reference to the original group, the solid area takes up less than a fifth of the area of the circle. In reference to the group which completed four semesters, the figure, 29.6, shows that over one-fourth achieved in this range.

The proportion in the 3.00 to 4.00 grade-point average range, represented by the solid area, can be traced through the four semesters. Note that, in reference to the original group, the proportion is fairly constant but in reference to the group completing each semester, the proportion increases. This reflects the observation that most of the students who do well the first semester not only continue to do well but also stay on in college. In like manner, the proportion in the 0.00 to 1.99 grade-point average range, represented by the lined areas, can be traced through four semesters. In this case, the proportion decreases in terms of both its reference to the original group and its reference to the group completing each semester. This reflects the observation that most of the students who earn less than a C average the first semester leave college early and that those who continue tend to raise their grades.

A comparison of the figures in the sections of the circle for the first semester with those in the circle for the fourth semester illustrates the significance of persistence or staying power. The group completing four semesters earned higher grades in the fourth semester than the larger, total group did in the first semester. This is partly the result of the departures of the less promising students and partly improvement in grades on the part of some of the students who continued.

The departure aspect of progress is shown by the percentages of the original number leaving at various times over a four semester period. Consider the figures shown below the circle for the first semester. The 1,924 men were selected on the basis of their having completed the first semester, and accordingly none left either before or during the first semester. Between the end of the first semester and the beginning of the second semester 5.1% left voluntarily and 2.2% was dropped. These percentages, 5.1% and 2.2%, are also shown in the set of figures under the circle for the second semester. Below the second semester circle it can be seen that an additional 3.0% left voluntarily during the semester and 0.4% was dropped during the semester. After the second semester and before the beginning of the third semester, another 8.4% left

voluntarily and another 11.2% was dropped. The figures after "total" show the cumulative effect of departures. Before the beginning of the third semester, 16.5% of the original group had left voluntarily and 13.8% had been dropped. Thus 30.3% of the original group had left before the third semester. Note that Summer Sessions are not accounted for. Departures shown as "after" include those that occur at any time between the close of one semester and the beginning of the next whether or not summer session attendance was involved.

The clear section in each circle shows the proportion of the original group not completing that semester. The actual percentage of the original group not completing the semester can be determined by adding together the four figures after the words "before" and "during". Thus these four figures under the circle for the first semester add up to zero, and there is no clear section in the circle. Under the circle for the second semester, the four figures (5.1, 3.0, 2.2, and 0.4) add up to 10.7%. The clear section represents this 10.7% of the original group. The figure of 10.7% is omitted from the circle so as not to confuse it with the figures in the shaded sections.

Very few students left during any semester. Voluntary departures occurred more than departures resulting from dropped actions. More departures occurred after the second semester than at any other time. Before the beginning of the fifth semester, normally the beginning of the junior year, 25.3% had left voluntarily and 21.4% had been dropped, thus making a total of 46.7% of the original group having left before the junior year. No account is given as to whether or not these students who left continued their college education elsewhere or whether or not they returned to the University of Wisconsin campus at some later time. The student who transfers from the College of Letters and Science to some other school or college within the University of Wisconsin after his first registration is nevertheless considered as being in the College of Letters and Science for the purposes of this analysis.

Charts II and III are designed exactly like Chart I but show the four semester progress of smaller groups selected from the 1,924 students represented in Chart I. These smaller groups were selected on the basis of previous academic achievement and scholastic ability.

Chart II shows the progress of the 266 students selected from the 1,924 on the basis of their having earned grades in the 3.00 to 4.00 grade-point average range the first semester and who also scored among the highest third of University of Wisconsin freshman men on the College Qualification Test. As would be expected, this more promising group not only earned higher grades but also a much larger proportion of them completed four consecutive semesters than was so for the total group of 1,924.

In order to illustrate the use of the prediction information in Chart II, assume that a student is being counseled during his second semester and that this student earned a grade-point average his first semester in the 3.00 to 4.00 range, and scored among the highest third of Wisconsin freshman men on the College Qualification Test. With this information, the counselor could make some reasonable predictions regarding the student's most likely progress. The probabilities are, for example, that this student will achieve in the same range as did the majority (68.6%) of the former group. The probabilities are even greater that he will earn a 2.50 grade-point average or above because 89.5% of this former group of like ability and achievement did so. It is almost certain that he will make at least a C average because 99.2% of this former group did so. Furthermore, it is unlikely that this student will leave during this second semester or even after its completion. None of the former group left during the semester and only 5.3% left after completing it. It is almost certain that he will not be dropped. In brief, the probabilities are that he will continue to do well and stay on in college. If, however, the counselor knows that this student is doing very poorly the second semester, then he will know that this is an unusual case and can begin to search for unusual factors which may be contributing either to his poor performance or his contemplated departure from the University.

Chart III shows the progress of another group of former students, a group of men who earned less than a C average the first semester and scored among the lowest third of entering Wisconsin freshman men on the College Qualification Test. Now consider the case of a student being counseled during his second semester, but this time it is a student who earned less than a C average the first semester and scored among the lowest third on the College Qualification Test. The probabilities are that this student will again earn less than a C average and that he will leave the University early in his college career; 65.3% of the former group attending the second semester earned less than a C average this second semester and by the end of the fourth semester 83.8% had left. Suppose, however, that this same student was being counseled during his fourth consecutive semester. He would then be among the minority who "survived." Most of the former group of "survivors" completed the fourth semester satisfactorily, and 20% earned a B average or higher.

Charts I, II, and III are three of 40 such charts which will be incorporated into a manual, "Progress Patterns for Counseling." Each of these 40 groups will be selected according to sex, previous academic achievement, and scholastic ability. Twenty of these charts will be based on high school achievement and ability and 20 on first semester college achievement and ability. This will permit use of the manual for both prospective and currently

enrolled students at the University of Wisconsin.

Counselors vary considerably in their use of predictive information. At one extreme are those who believe that predictive information interferes with counseling relationships. At the other extreme are those who focus on objective data with little concern for feelings and attitudes. Most counselors recognize the significance of both the feelings and the objective data. Drake and Oetting[3] have described this middle ground in their discussion of the use of psycho-metric data in counseling. The manual, "Progress Patterns for Counseling," is being developed for use in counseling situations. This manual will, of course, be limited to use with prospective or currently enrolled students of the College of Letters and Science, University of Wisconsin. Counselors at Wisconsin will be introduced to this manual for use on a trial basis. Experience with it may result in some substantial revisions. In the meantime, others may find that progress patterns of similar design may be helpful to counselors at their institutions.

Summary

Scholastic progress patterns, developed from the records of former students, could be used as predictive information in counseling with prospective or currently enrolled students of that college. These patterns were developed from two basic premises. One is that previous academic achievement and scholastic ability are significantly related to progress. The other is that the scholastic progress made by prospective or currently enrolled students of a college will be essentially the same as progress made by previously enrolled students of that college. Evidence to support these premises is in published reports and in the records of former students at the University of Wisconsin. Examples of progress patterns for three different groups show how progress can be described and used in counseling situations. Patterns of this kind will be used in some counseling situations at the University of Wisconsin on a trial basis.

FOOTNOTES

1. R. M. Travers, "Significant Research on the Prediction of Academic Success," in Wilma T. Donahue (Ed), The Measurement of Student Adjustment and Achievement, (Ann Arbor: University of Michigan, 1949), pp. 147-190.
2. H. F. Garret, "A Review and Interpretation of Investigations of Factors Related to Scholastic Success in Colleges of Arts and Sciences and Teachers Colleges," Journal of Experimental Education, XVIII (1949), pp. 91-138.
3. L. E. Drake and E. R. Oetting, An MMPI Codebook for Counselors, (Minneapolis: University of Minnesota Press, 1959), 140 pp.

TABLE I

DEPARTURES BETWEEN END OF FIRST SEMESTER OF FRESHMAN YEAR AND BEGINNING OF FIFTH CONSECUTIVE SEMESTER ACCORDING TO SCHOLASTIC PROMISE AND SEX, FOR UNDERGRADUATES COLLEGE OF LETTERS AND SCIENCE, ENTERING IN SEPTEMBER, 1958 OR 1959

Scholastic Promise	Departures as Percentages of N					
	Voluntary		Dropped		Total	
	Men	Women	Men	Women	Men	Women
High Scholastic Promise *	16.9	38.0	1.2	0.0	18.2	38.0
Low Scholastic Promise **	29.2	41.2	54.6	45.0	83.8	86.2
TOTAL ***	25.3	41.6	21.4	12.5	46.6	54.1

* High scholastic promise: students who earned a 3.00 G.P.A. or above the first semester and ranked among the highest third of entering University of Wisconsin freshmen of their sex on the College Qualification Test. Number of men: 266; number of women: 332.

** Low scholastic promise: students who earned less than a 2.00 G.P.A. the first semester and ranked among the lowest third of University of Wisconsin freshmen of their sex on the College Qualification Test. Number of men: 408; number of women: 340.

***Total: all new freshmen of the College of Letters and Science who completed the first semester beginning in September 1958 or September 1959. Number of men: 1,924; number of women: 2,022.

TABLE II

*GRADE-POINT AVERAGES OVER TEN-YEAR PERIOD FOR UNDERGRADUATE MEN ENROLLED SECOND SEMESTER, COLLEGE OF LETTERS AND SCIENCE**

Year	Freshmen		Sophomores		Juniors		Seniors		Total	
	No.	GPA	No.	GPA	No.	GPA	No.	GPA	No.	GPA
1952	765	2.28	720	2.55	714	2.71	825	2.79	3024	2.58
1953	980	2.31	714	2.62	609	2.75	759	2.84	3062	2.60
1954	923	2.29	906	2.64	576	2.78	647	2.86	3052	2.61
1955	860	2.29	919	2.54	719	2.76	649	2.81	3147	2.58
1956	894	2.27	908	2.57	688	2.70	775	2.86	3265	2.58
1957	827	2.28	894	2.50	657	2.63	775	2.78	3153	2.54
1958	825	2.32	827	2.51	629	2.63	742	2.81	3023	2.56
1959	989	2.22	846	2.50	626	2.70	707	2.79	3168	2.52
1960	1144	2.28	960	2.51	600	2.60	696	2.80	3400	2.51
1961	1412	2.30	1094	2.54	784	2.67	698	2.83	3988	2.53
GPA Range	2.22-2.32		2.50-2.64		2.60-2.78		2.78-2.86		2.51-2.61	

*Source: Undergraduate Student Scholarship Records, Office of the Registrar, The University of Wisconsin, Madison.

TABLE III

DEPARTURES DURING FIRST FOUR CONSECUTIVE SEMESTERS, SEPTEMBER 1958 AND SEPTEMBER 1959 ENTERING FRESHMAN MEN, COLLEGE OF LETTERS AND SCIENCE

When Departures Occurred	Departures as Percentages of N*					
	Voluntary		Dropped		Total	
	1958	1959	1958	1959	1958	1959
After First Semester**	5.7	4.5	2.0	2.5	7.7	7.0
After Second Semester	12.6	10.5	12.5	10.6	25.1	21.1
After Third Semester	3.7	3.4	5.3	5.3	9.0	8.7
After Fourth Semester	4.9	5.6	2.8	2.1	7.7	7.7
TOTAL	26.9	24.0	22.6	20.5	49.5	44.5

* N for 1958 group is 874; N for 1959 group is 1050.

**After: any time from the beginning of one semester to the beginning of the next. This includes both those who left during a semester and those who left between the end of one semester and the beginning of the next. Summer Session attendance is not accounted for.

CHART I

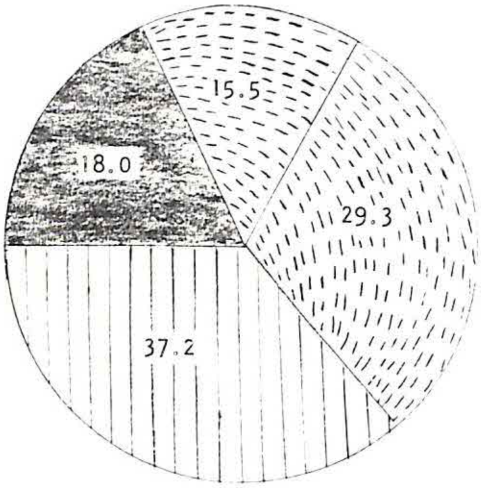
PROGRESS MADE DURING FIRST FOUR CONSECUTIVE SEMESTERS BY STUDENTS WHO FIRST ENTERED THE COLLEGE OF LETTERS AND SCIENCE UNIVERSITY OF WISCONSIN, SEPTEMBER 1958 OR 1959 AS NEW FRESHMEN

KEY: Areas: solid, 3.00 – 4.00 GPA; dotted, 2.00 – 2.99 GPA (line at 2.50); lines, 0.00 – 1.99 GPA; clear, those not completing that semester.

Numbers: Those in areas are percentages of number completing that semester. Those for departures are percentages of number of original group.

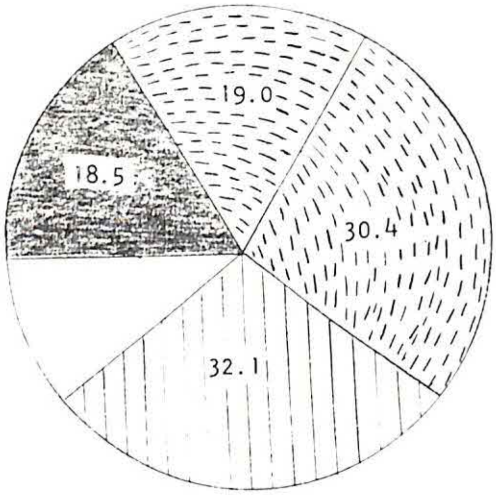
ORIGINAL GROUP: 1,924 men; all freshmen completing first semester.

First Semester



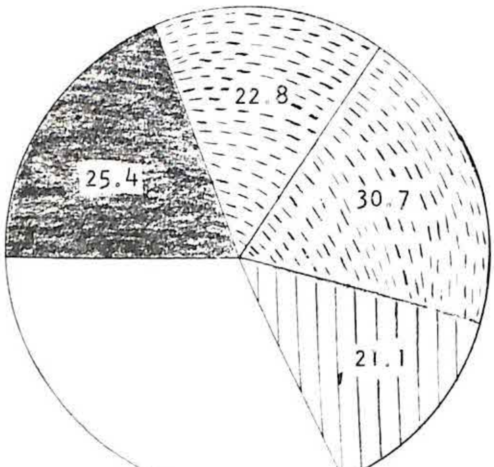
Departures:	Vol.	Drop
Before	0.0	0.0
During	0.0	0.0
After	5.1	2.2
Total	5.1	2.2

Second Semester



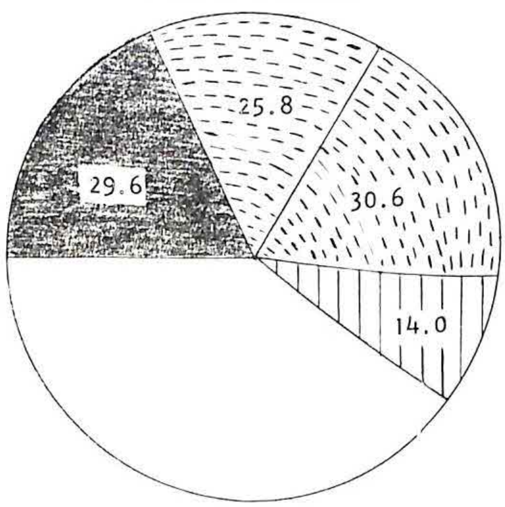
Departures.	Vol.	Drop
Before	5.1	2.2
During	3.0	0.4
After	8.4	11.2
Total	16.5	13.8

Third Semester



Departures:	Vol.	Drop
Before	16.5	13.8
During	1.5	0.2
After	2.0	5.0
Total	20.0	19.0

Fourth Semester



Departures:	Vol.	Drop
Before	20.0	19.0
During	1.5	0.0
After	3.8	2.4
Total	25.3	21.4

CHART II

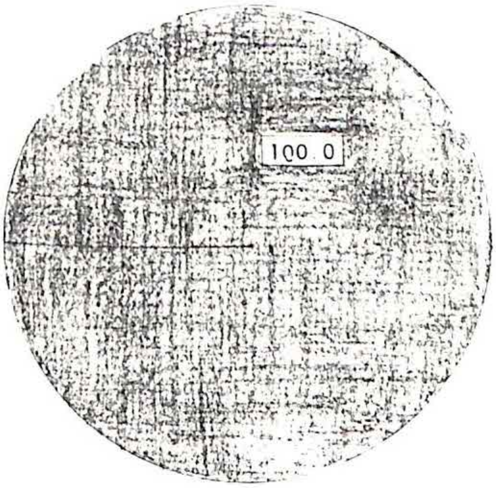
PROGRESS MADE DURING FIRST FOUR CONSECUTIVE SEMESTERS BY STUDENTS WHO FIRST ENTERED THE COLLEGE OF LETTERS AND SCIENCE UNIVERSITY OF WISCONSIN, SEPTEMBER 1958 OR 1959 AS NEW FRESHMEN

KEY: Areas: solid, 3.00 – 4.00 GPA; dotted, 2.00 – 2.99 GPA (line at 2.50); lines, 0.00 – 1.99 GPA; clear, those not completing that semester.

Numbers: Those in areas are percentages of number completing that semester. Those for departures are percentages of number of original group.

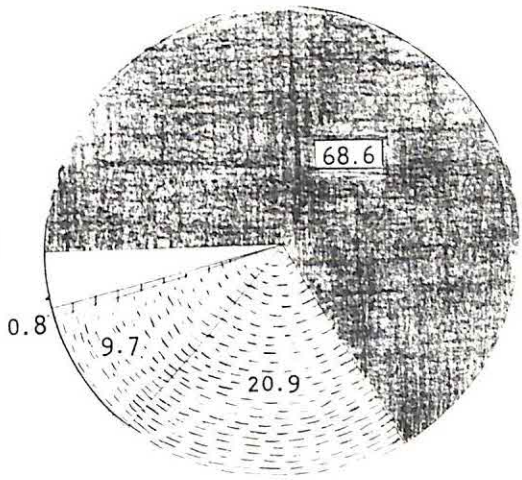
ORIGINAL GROUP: 266 men; 1st sem. GPA 3.00 – 4.00; CQT upper third.

First Semester



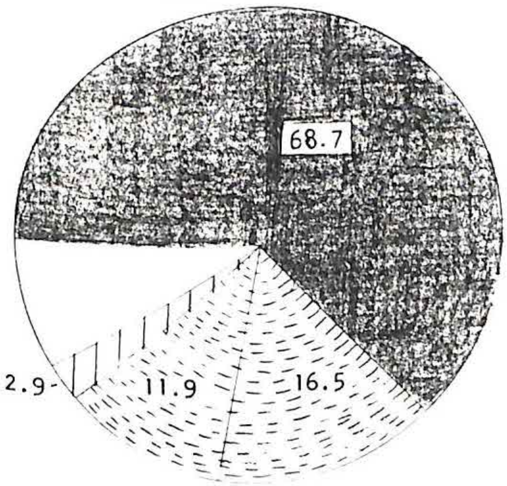
Departures:	Vol.	Drop
Before	0.0	0.0
During	0.0	0.0
After	3.0	0.0
Total	3.0	0.0

Second Semester



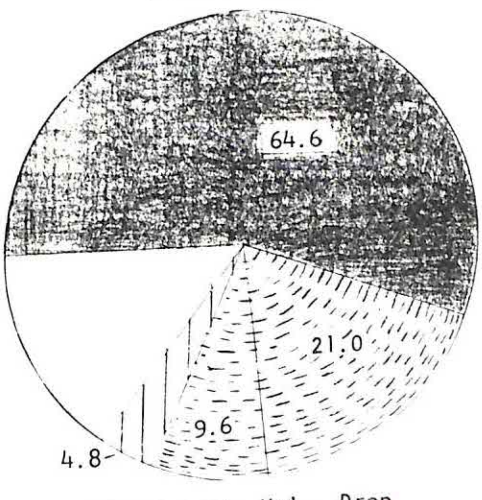
Departures:	Vol.	Drop
Before	3.0	0.0
During	0.0	0.0
After	5.3	0.0
Total	8.3	0.0

Third Semester



Departures:	Vol.	Drop
Before	8.3	0.0
During	0.4	0.0
After	3.3	0.4
Total	12.0	0.4

Fourth Semester



Departures:	Vol.	Drop
Before	12.0	0.4
During	1.5	0.0
After	3.4	0.8
Total	16.9	1.2

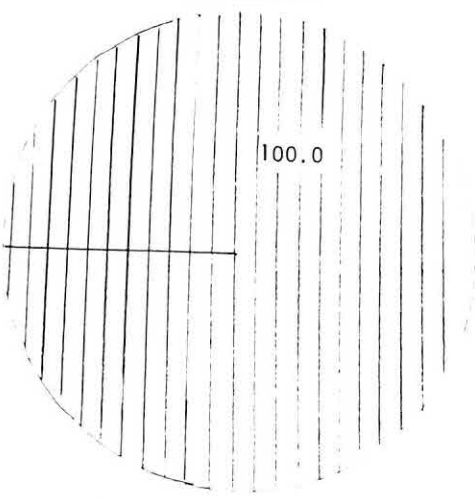
CHART III

PROGRESS MADE DURING FIRST FOUR CONSECUTIVE SEMESTERS BY STUDENTS WHO FIRST ENTERED THE COLLEGE OF LETTERS AND SCIENCE UNIVERSITY OF WISCONSIN, SEPTEMBER 1958 OR 1959 AS NEW FRESHMEN

KEY: Areas: solid, 3.00 – 4.00 GPA; dotted, 2.00 – 2.99 GPA (line at 2.50); lines, 0.00 – 1.99 GPA; clear, those not completing that semester.
Numbers: Those in areas are percentages of number completing that semester. Those for departures are percentages of number of original group.

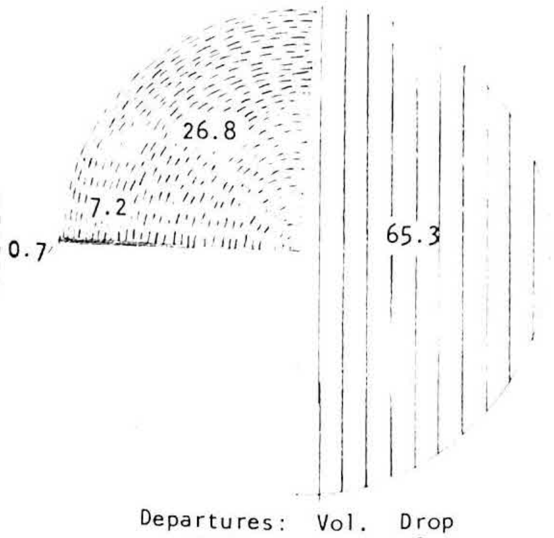
ORIGINIAL GROUP: 408 men; 1st sem. GPA 0.00 – 1.99; CQT lower third.

First Semester



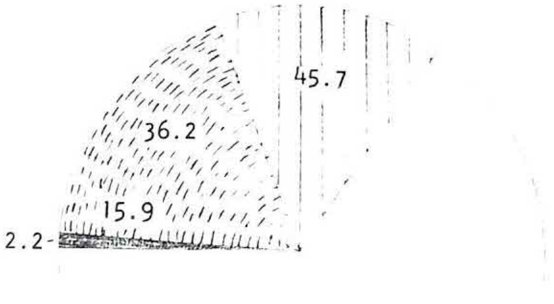
Departures:	Vol.	Drop
Before	0.0	0.0
During	0.0	0.0
After	9.1	8.3
Total	9.1	8.3

Second Semester



Departures:	Vol.	Drop
Before	9.1	8.3
During	6.3	1.3
After	7.9	30.8
Total	23.3	40.4

Third Semester



Departures:	Vol.	Drop
Before	23.3	40.4
During	2.0	0.5
After	1.7	11.0
Total	27.0	51.9

Fourth Semester

Departures:	Vol.	Drop
Before	27.0	51.9
During	0.3	0.0
After	1.9	2.7
Total	29.2	54.6

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

PREDICTING MEDICAL SCHOOL SUCCESS: A TEN-YEAR STUDY

ROGER GRATWICK, JAMES DRASGOW, BRUCE STOCKIN
University of Buffalo

THE PROBLEM of predicting success in medical school has received considerable attention particularly during recent years. Gottheil and Michael(4) have made the most extensive study of studies with their review of the research on the topic. They concluded that psychological tests are the most promising means for predicting which students will succeed in medical training.

The present authors feel this conclusion should be carefully examined lest the implications and certain fundamental assumptions be clouded over or forgotten. A primary concern facing any educator who is responsible for selecting students for any training program is the ability to estimate as accurately, and as early as possible, the probability that a candidate will succeed or fail. Furthermore, devices used to select the potentially successful students need to meet the criteria of economy, practicality, and reliability of differentiating "successes" from probable "failures."

The results of 10 years' work at the University of Buffalo on the prediction issue have convinced the authors that a workable solution to this problem may indeed be in the direction suggested by Gottheil and Michael, but may also be more readily available than previously believed. Our findings and their implications are presented in the hope that they may stimulate others toward new viewpoints on the issues.

Phase I: 1949

In this initial pilot study the relationship of pre-med grades in sciences (biology, physics, and inorganic and organic chemistry) with grades received during the freshman year in medical school was examined. The pilot research sample consisted of 54 students who had done their pre-medical work at the University of Buffalo's College of Arts and Sciences and then went on into the University's Medical School. The correlation found between the basic science grades and freshman medical school grades was .75, which indicated a very intimate and respectable relationship.

Since science course grades are usually easily available, this finding suggested that we may have discovered a potentially very practical and useful predictor variable which can be used for cross-validation at other schools.

Scores on the Medical College Admissions Test were then examined to determine their relationship to first-year grades in medical school. On three of the test's four scales ("Q," "Verbal," and "Social Organization") the correlations were very low (.25, .03, and .28 respectively). All of this indicated practically no usable relationship for individual predictions. With the fourth scale, "Science," a correlation of .42 was obtained. Although this correlation does suggest a relationship, it does not approach that of the previously found relationship with grades from required science courses; nevertheless, we wondered if it might not be possible to combine science course grades with the science sub-test score to produce an improved predictor. The statistical work revealed that the test scores were highly correlated with grades and did not add significantly to produce a better predictor.

These findings tentatively suggested that perhaps one of the most easily available and "objective" measures for prediction was also the best one. However, educational and psychological information made us reluctant to accept such evidence without considering other variables that might increase predictive accuracy.

Phase II: 1952

The second phase focused on the subjective evaluation made by the Pre-Medical Appraisal Committee of each candidate. For many years it has been the practice at the University of Buffalo to obtain subjective ratings by various personnel of all pre-medical students. This work is done by a committee of four: three from the departments of Chemistry, Physics, and Biology plus a Chairman from the Office of the Dean of Students. Each of the three science departments has a card rating system whereby all medical school candidates are rated by their lecture professors and laboratory assistants. The committee chairman interviews each student and also has access to the opinions of counselors, administrative officers in the Student Union and Head Residents in the dormitories.

During his pre-med work, each student is rated

on a four-point scale in the following four areas: general impression and personality, emotional adjustment and stability, creative experimental interest, and industry. At the close of his pre-medical studies, an appraisal is made of each candidate in a meeting of the entire committee and each candidate is rated for medical school acceptance as either strongly recommended, recommended, average, or not recommended.

In this investigation, we compared the subjective evaluations of candidates with their success (defined as average or better course grades) in first-year medical school. The sample consisted of 81 students who had been evaluated by the science departments and the Appraisal Committee. The findings revealed a relationship between the overall committee appraisals and success in medical school. Stated in the form of prediction probabilities, the chances for success in the first year of medical school of those candidates rated as "strongly recommended" or "recommended" were eight to one in their favor. The probability of success for those candidates rated as "average" was three to one. None of the students given a "not recommended" rating was accepted for medical school.

Although the overall Appraisal Committee ratings were related to medical school grades, individual department ratings were not. In other words, the subjective personality evaluations made by the individual observers bore no significant relationship to later success or failure in medical school. This suggested that the Appraisal Committee was using something which was not being used by the individual observers, or which was not contained in the "personality" data.

In searching for an explanation of these findings, correlations were run between the Appraisal Committee's recommendations and many variables available to them, e.g., number of high-school activities, regents average, fifth in high school, grade-point average in pre-med studies, educational level of parents, service record, college aptitude test scores, anxiety test index, etc. The surprising finding was that the only respectable correlations obtained were between committee recommendation and 1) regents average (.69), and 2) college grade-point average (.68). In spite of the fact that grades were not the committee's primary concern, it was evident that a student's committee approval was more closely associated with his marks in high school and pre-med studies than with any other variable. Since Phase I three years earlier had shown substantially the same thing in principle, by this time we were really wondering if grades were the only variable that was related to medical school success. Hence, attention was turned to a more thorough investigation of these variables.

Phase III: 1958

For this part of the project, the pre-medical and medical school records were accumulated on 240 medical students who had done their pre-med work at the University of Buffalo and were at that time beyond their first year of medical school. These students were then divided into three groups: 1) above average med school grades, 2) average grades, and 3) below average. Next, all background variables on which data were available were examined for their relationship to medical grades; these included high-school activities, number and kinds of hobbies, father's and mother's occupational level, father's occupation, foreign language background, position in the family, number of siblings, family income, and high-school job history.

There were no significant differences among the three groups on any of these background variables. It apparently did not matter what the socio-economic background or extra-curricular history of the student had been because his marks in medical school might be randomly poor, average, or excellent. Consequently, the search for background factors related to degree of success within medical school revealed no relevant highly correlated variables.

At this point it was decided to study the possibility of moving predictions of success or failure forward in time to the beginning of pre-med work. Our most recent investigation, therefore, gathered data on the feasibility and possible value of earlier prediction.

Phase IV: 1960

This phase of the work was designed to focus on the entering undergraduates aspiring to be physicians. Since the prediction problem is similar in principle for dental as well as medical candidates, it was decided that both of these specialities could be used and compared. Consequently, a sample of 101 students who had enrolled in a pre-medical or pre-dental curriculum upon entering college was selected.

Fifty-nine of these students had classified themselves in 1954 as pre-medical and 42 listed themselves as pre-dental. Each was followed up in 1959, which was five years after entering; ample time therefore was allowed for practically all of them to have achieved entrance to either professional school.

The findings were quite unexpected. Only 10 of the pre-medical students achieved the goal of admission to medical school. The situation with the dental students was somewhat similar; only 11 were admitted to dental school. Only half of the original group of students achieved any kind of educational goal; 45 percent discontinued their education completely, and the other five percent were still working on their bachelor's degree.

Implications and Conclusions

What have all these studies shown? Several

things seem evident. First, the search for complicated psychological measures to predict medical school success may be a waste of time and resources; the psychology literature is loaded with studies attempting to use the Rorschach, Thematic Apperception Test, and various scales of the Minnesota Multiphasic Personality Inventory as well as other psychological tests to predict academic achievement, but as yet none has been reported to have a validity coefficient approaching that of the science course grades. From our experiences it appears that grades are more closely related to performance in medical school courses than are less objective and less easily obtained indices.

Grades in the required science courses appear to indicate most accurately what may be expected from a student in his medical school work. However, the finding that high-school regents grades are highly related to recommendations for medical school suggests that perhaps it isn't necessary to await college grades to make a very good prediction of a student's chances of success in medical school.

The fact that so few students who plan to go to medical school actually get there points up the need for early evaluation and appropriate guidance. But what seems to occur is that everybody views the college pre-medical studies as an academic "proving ground" to demonstrate the student's fitness for medical training. Operating from this framework, the Appraisal Committee provides for close observation of the candidate's progress during three or four years of college work and collects detailed ratings of various personality characteristics—all of which bear little or no relationship to eventual performance in medical school. Nevertheless, the Committee's overall evaluation made at the threshold of medical school has been shown to relate very well to performance in medical training. Why, we might ask, need the committee wait three or four years to make an evaluation that, while apparently valid, is based mainly on something available at a much earlier date? Why not appraise candidates before they have spent three or four years in preparation for a goal that three-quarters of them never achieve?

At this point, it may appear we are taking a naive approach to a complex problem; that we are saying the only thing of importance in selecting medical students is academic ability. Let us hasten to correct such a possible misinterpretation by suggesting what we think is a sensible solution to the prediction problem.

A well-known, highly regarded, and repeatedly found principle underlies our proposed solution. It is that the best indication of what a person will do in the future is what he has done in the past. For instance, if an individual has shown academic achievement in high school, he can be expected to show similar performance in college or in a graduate school. This principle is also one of the cornerstones for theorizing in the area of vocational development and counseling today. It is held to be valid for not just academic performance but for a person's entire "life style" as well and is used in psychodiagnostic work frequently.

There is considerable research evidence (1,3,6) to indicate that when intelligence (or academic ability) is held constant, the one factor which discriminates success and failure best is test-measured interest.

If our principle is valid and the research reliable, why must we push for complicated psychological instruments and why must we await the results of pre-med college work before making our predictions of success or failure for medical school? Rather, at the beginning of a student's college career, make an evaluation of his probable chances of success in medical school on the basis of his academic performance to date and his measured interest in the medical field. If the probability of success is high, encourage him to pursue his goal; if the probability is low, provide appropriate guidance immediately instead of three or four years later when the lack of interest or ability becomes painfully evident to the student as well as the school.

To the present authors it would seem that a pre-college evaluation on the ability-interest dimensions would be both practical and adequate. Skillful use of the Strong Vocational Interest Blank, for example, coupled with knowledge of the individual's academic potential shown by his past achievement, and a careful extrapolation of probable future behavior would yield a prediction of medical school success at a time when it would be of most value to the student and to those investing in his education. This shift in perspective combined with a clear awareness of the goal of prediction could clear the current fog of confusion and doubt.

REFERENCES

1. Berdie, R. F. "Aptitude, Achievement, Interest, and Personality Tests: A Longitudinal Comparison," Journal of Applied Psychology, XXXIX (1955), pp. 103-14.
2. Combs, A. W. "Non-directive Techniques and Vocational Counseling," Occupations, XXV (1946), pp. 261-67.
3. Darley, J. G., and Hagenah, T. Vocational Interest Measurement (Minneapolis: University of Minnesota Press, 1955).
4. Gottheil, E., and Michael, C. "Predictor Variables Employed in Research on the Selection of Medical Students," Journal of Medical Education, XXXII (1957), pp. 131-47.
5. Herzberg, F. The Motivation to Work (New York: Wiley and Son, 1959).
6. Kelly, E. L., and Fisk, D. W. "The Prediction of Performance in Clinical Psychology," (Ann Arbor, Mich.: University of Michigan Press, 1951).
7. Super, D. E. The Dynamics of Vocational Adjustment (New York: Harper Brothers, 1942).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

REASONS FOR ACADEMIC FAILURE

ADELE M. MILLER
Brooklyn College

"I LIKED COLLEGE--too bad I never bothered to open a book." That comment was made on a questionnaire returned to the College by a student who had been dropped for poor scholarship. It seems to symbolize the tragic waste of time and talent that is the annual toll of dismissals.

Why do students flounder and eventually sink in the academic seas? This is the eternal question that besets so many students, counselors, and administrators.

Background

Brooklyn College, now part of the newly organized City University of New York, is a municipal college supported by tax funds. The rather homogenous student body is, with few exceptions, drawn from the city, and more particularly, from the borough of Brooklyn. Most of the students are graduates of the New York City secondary schools and are admitted to Brooklyn College on the basis of their high school averages or a combined total of their high school average and college entrance examination scores.

Brooklyn College expects its students to maintain a "C" average. Those students whose records fall below "C" are warned and may be dismissed if they show no marked improvement.

Dismissal statistics of the College of Liberal Arts and Sciences, Table I, may be viewed against this general background of the standards expected of students. The statistics show that a higher proportion of men than of women have been dismissed. Furthermore, with two exceptions, more dismissals have been made in the sophomore year than any other year (see Table II). It has been the policy of the College to be somewhat more lenient with freshmen since the first year entails so many adjustments.

Of course, these statistics do not present the total picture of failure to meet academic standards. Additional students have been dropped at mid-year although these numbers were much lower than those in June. Also, many students dropped out of their own accord. Data of students who entered Brooklyn College in February 1953 and who withdrew from college showed that 51.2 percent had unsatisfactory records.[1] It is of interest to note that only 22 percent of the men and three percent of the women who applied to leave College indicated that they were dropping out because of lack of interest in or ability to cope with their studies.[2] Professor Eilenberg's study has also pointed out that of 631 students queried, 432 had some difficulty in making an academic adjustment to college.

The Purpose of the Studies

Even though the total number of dismissals has been small compared to the total enrollment, the College has felt an obligation to each student admitted. It was felt that with more detailed knowledge of the reasons for failure, perhaps some help could be provided to students in danger of dismissal. With this in mind, several studies were undertaken. The aim of the studies to be described was to try to determine the reasons for academic failure and to take appropriate steps in the light of the information gathered.

The Procedures

Studies were conducted using different sources of information. In one study, 45 students who had improved their records sufficiently so as to be free of academic warning were questioned as to their original inability to do successful college work. Eleven "reasons" for unsuccessful work were listed (Table III), and the students were asked to rank those reasons in order of importance as they felt they applied to themselves. The problems that confront students at a resident college, i.e., living away from home for the first time, do not exist here. On the other hand, some students find living at home at the time when they seek to be independent, very trying.

During the course of the semester in which a student receives an academic warning, his counselor calls him in with a view toward both helping the student and helping the administrative officers of the College by reporting on the student. An analysis of the counselor's reports on 331 such students, subsequently dismissed, indicates the following reasons for failure to do satisfactory work (Table V). Where more than one reason was cited, only the one that seemed most important was tallied.

---

1 Jeanette H. Eilenberg, Brooklyn College, Class Portrait 1953-1957 (New York: Brooklyn College, 1959), p. 25.

2 Ibid.

TABLE I

COMPARISON OF DISMISSALS WITH ENROLLMENTS IN THE SPRING OF EACH YEAR, 1952-61
COLLEGE OF LIBERAL ARTS AND SCIENCES

Year	Men			Women			Total		
	Total No. Enrolled	Dismissed No.	Dismissed %	Total No. Enrolled	Dismissed No.	Dismissed %	Total No. Enrolled	Dismissed No.	Dismissed %
1952	3790	155	4.1	4219	40	0.9	8009	195	2.4
1953	3561	108	3.0	4231	29	0.7	7792	137	1.7
1954	3723	177	4.7	4422	47	1.1	8145	224	2.8
1955	3512	175	4.9	4350	65	1.5	7862	240	3.1
1956	3413	125	3.6	4260	42	1.0	7673	167	2.2
1957	3453	89	2.5	4315	38	0.9	7768	127	1.6
1958	3211	126	3.9	4299	59	1.4	7510	185	2.5
1959	3272	100	3.1	4504	46	1.0	7776	146	1.9
1960	3383	130	3.5	4687	43	0.9	8070	173	2.1
1961	3562	123	3.5	4957	60	1.2	8519	183	2.1

TABLE II

DISMISSALS BY CLASS AND SEX IN JUNE OF EACH YEAR, 1952-61
COLLEGE OF LIBERAL ARTS AND SCIENCES

Year	Freshmen			Sophomores			Juniors			Seniors		
	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
1952	63	18	81	55	15	70	27	7	34	10	--	10
1953	32	2	34	48	19	67	20	6	26	8	2	10
1954	102	21	123	48	19	67	24	5	29	3	2	5
1955	51	22	73	95	39	134	23	2	25	6	2	8
1956	30	7	37	58	28	86	26	6	32	11	1	12
1957	38	6	44	39	21	60	6	9	15	6	2	8
1958	37	10	47	62	36	98	26	12	38	1	1	2
1959	44	10	54	30	27	57	19	8	27	7	1	8
1960	40	7	47	57	25	82	27	10	37	6	1	7
1961	24	7	31	54	43	97	31	8	39	14	2	16

TABLE III

SURVEY OF STUDENT OPINION ON REASONS FOR FAILURE TO DO SATISFACTORY WORK

Reason	Frequency of Mention in Order of Importance										
	1	2	3	4	5	6	7	8	9	10	11
Poor Study Habits	12	10	6	1	1	--	1	--	--	--	--
Insufficient Study	9	5	5	6	--	--	1	--	--	--	--
Lack of Interest in Subjects	2	5	3	6	3	4	--	--	1	--	--
Emotional Problems	6	3	3	1	2	1	1	1	2	--	--
No Vocational Goal	6	3	2	2	2	1	--	--	--	--	--
Home Circumstances	2	5	--	--	3	--	3	--	--	--	--
Insufficient H. S. Preparation	2	2	3	--	2	1	1	1	--	2	--
Inadequate Study Facilities	1	2	2	3	4	--	--	--	--	--	--
Employment	2	--	3	2	--	--	1	2	--	--	1
Co-curricular Activities	--	2	1	1	2	1	--	2	--	--	2
Illness	3	1	--	--	1	1	--	2	1	1	1

TABLE IV

WEIGHTED RANK VALUE OF REASONS LISTED*

Reason	Value
Poor Study Habits	306
Insufficient Study	247
Lack of Interest	195
Emotional Problems	166
No Vocational Goal	150
Home Circumstances	108
Insufficient H. S. Preparation	102
Improper Study Facilities	101
Employment	79
Co-curricular Activities	72
Illness	70

* Each time a reason was listed as number 1 it was weighted 11 times; each time it was listed as number 2 it was weighted 10; etc.

TABLE V

REASONS FOR FAILURE AS INDICATED BY COUNSELING REPORTS

Reason	No. of Students Reported
Lack of Application--Poor Study Habits	80
Lack of Goals--Poor Motivation	47
Lack of Ability	44
Wrong Major Field or Course Choice	34
Employment or Financial Problems	33
Family Problems	32
Psychological-personal Problems	29
Poor Adjustment to College	17
Co-curricular Activities	8
Poor Health	7
Total	331

Recently a questionnaire was sent to 31 students who had been dropped in February of 1960. Some of the respondents had comments to make in addition to the replies to the specific questions. Some of the comments were as follows: "My trouble was that nothing bothered me. I liked College--too bad I never bothered to open a book"; "My low scholarship status was not due to any lack of ability to do the work. I took a very immature (and irresponsible) attitude toward my studies... My records show I got several A's and B's in certain subjects. These were subjects I liked and studied (and learned with a minimum of effort). The other subjects I just did not care for, and so I did not devote much time to them. Moreover, I gave too much of myself to other activities that were not important."

From the above, it seems that students can give quite dispassionate appraisals of themselves.

Implications for the College

(1) At the present time the criteria for admission are being examined in the light of student performance to determine whether any adjustment should be made in the weight given to the entrance scores. In view of the tax-supported nature of the institution, it is unlikely that any subjective factors can be given weight over and beyond the grade achievement in the secondary school and the entrance examination scores.

(2) In the very first days of their college experience, it would seem desirable that the students have explained to them the new study habits, both in quantitative and qualitative measures, that are necessary for them to succeed.

(3) The College has also established a Basic Skills Center where, at a nominal charge, students may take a course aimed at improvement of their reading and study skills.

(4) The Personal Counseling Program, a special subdivision of the counseling facilities at the College, seeks to help the student, through professional guidance, to work through his personal difficulties.

(5) If the student does not meet the academic standards of the College after effort has been made to help him perform at the level of his potential, dismissal may be a more meaningful educational experience. Sometimes the experience is painful, but for many it is the first time they have had to take a long, hard look at what they were doing. For some, dismissal may be a relief from the obligation of doing something which they themselves did not want in the first place.

At Brooklyn College, it has been the policy to give to any dismissed student, who has been away for a year and who seems to have greater maturity and a new seriousness of purpose, a "second chance" as a part-time, tuition-paying student. If he does well as a part-time student, he may be restored to full matriculation. Strange as it may seem, some of these students applying for readmission have indicated that their dismissal was the best thing that ever happened to them. The following is a quotation from an application for readmission:

> Although I at times realized that the day would eventually come when I would be asked to leave, I still could not fully comprehend that such a thing could and would happen to me and so it did. May I say that now I am truly sorry that this did not happen two years previous.....
>
> ....I have had plenty of time to see as well as to make comparisons between these two worlds, but most of all I have had time to think; time to reflect upon what has transpired in the past; what was, what will be, but most of all what can be for me.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

DEVELOPING UNIFORM DEPARTMENTAL GRADING STANDARDS IN A UNIVERSITY

O. F. ANDERHALTER
Saint Louis University

UNIVERSITY ADMINISTRATORS are frequently concerned with the problem of apparent differences in grading standards from department to department in the various schools of a university. Students may be given grades by one department differing considerably as a group from those given the same students during the same semester by another department. Efforts to remedy the situation through written instruction, faculty meetings, and other similar methods, are generally unsuccessful.

An attempt has been made by the Bureau of Institutional Research of Saint Louis University to develop a method pointed toward solving the problem. In essence, the method consists of a procedure for the analyses of data which quantifies the apparent inequities in such grading standards, and which does so in a manner understandable to the average teacher. Since the problem is a continuing one, it is essential that the procedures be repeated annually.

Quantifying Departmental Inequities in Grading Standards

The technique adopted for the analyses of data is basically a simplified and "spelled-out" version of analyses of covariance. Using a freshman class in the College of Arts and Sciences at Saint Louis University for illustrative purposes, the technique consists of the following:

(1) Following the completion of the school year, the credit-point-average (CPA) is determined for all students independent of departments from which grades were obtained.

(2) Measures available at the time students were accepted by the University are then correlated with end-of-year CPA's and with conventional methods used for selection of predictors to be used in multiple correlation and regression analyses. For the example being used, a combination of high school rank and entrance examination scores yielded a multiple coefficient of correlation of .682.

(3) The multiple regression equation is then used to "predict" the CPA expected for each incoming freshman student. The predicted CPA is referred to as the "potential index" for each student.

(4) For each department this potential index is identified for all students receiving a grade from the department. Thus, should 50 students have taken a course in English, the potential index for each of these 50 students would be determined.

(5) For each department the mean potential index is compared with the comparable mean CPA for grades assigned. Thus, should an English department assign 50 grades, giving an average 2.5 CPA, this 2.5 would be compared with the average potential index of the 50 students involved. Where the potential is lower than the grades assigned, there is some indication that the department is "too liberal" in its grading practices; where the potential is higher than the grades assigned, there is evidence that the department is "too tough" in its grading practices.

The first four columns of Table I illustrate the results from these types of analyses for one freshman class of Saint Louis University. In addition, within each department, the potential index was correlated with grades received. A high degree of relationship indicates that the "better" students were receiving the better grades given by the department.

Use of Results in Modifying Grading Standards

After the results become available, apparent inequities can more adequately be explained to the department administrators. It becomes apparent, for example, that department number 3 is grading "too tough." Grades assigned are considerably below the potential of individuals as determined through their overall performance in all course work. Department number 13 is giving grades which on the average are almost two letter grades higher than grades given the same students by

TABLE I

EVALUATION OF DEPARTMENTAL GRADING PROCEDURES

Coded Department	(1) Number of Students	(2) Mean Credit-Point Average	(3) Mean Potential Index	(4) Departmental Grading Index*
1	116	2.28	1.80	.48
2	169	2.13	2.06	.07
3	57	1.49	2.25	-.76
4	154	2.17	1.78	.39
5	111	2.18	1.81	.37
6	45	2.42	2.14	.28
7	437	2.25	2.03	.22
8	39	2.51	2.24	.27
9	100	2.03	2.38	-.35
10	87	2.10	2.00	.10
11	20	2.55	2.28	.27
12	354	2.19	2.08	.11
13	17	3.11	1.51	1.60
14	222	2.85	2.18	.67
15	46	2.15	2.24	-.09
16	320	2.13	2.08	.05
17	23	1.56	1.50	.06
18	105	2.19	1.81	.38
19	20	2.00	1.36	.64
20	47	1.95	2.13	-.18

* Difference between Mean Credit-Point-Average and Mean Potential Index.

other departments. In both of these examples the correlations were relatively lower than most of the other correlations. Not only was the "level" of the grading questionable, but also the relationship between grades and potential was relatively lower.

Other results can be interpreted to departmental administrators in a similar manner. Faculty in the departments can then be alerted to the results, and an effort can be made to "equalize" the standards.

Summary

The procedures outlined in this paper have resulted in a considerable improvement in equalization of grading standards by the various departments. While the method does not take into account differential aptitudes of individuals — which may influence interpretation of the results provided — it is likely that such factors would tend to be balanced as the number of students involved increases.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

THREE ERROR SOURCES IN COLLEGE GRADING

BERNARD C. KIRBY
San Diego State College

THE PURPOSE of this study is to investigate certain distortions of the college grading process, distortions which have in common one element: they are all considered to be matters of luck. The attempt is to isolate certain kinds of luck and then measure very roughly the extent to which they explain variance in grade-point average.

The term luck, or its synonyms, is seldom defined precisely; called chance, it is the subject of whole volumes of probability theory, but it still is used colloquially more than scientifically. Here it refers to classes of events or more generally, factors, which are 1) either unknown or unknowable, or 2) either uncontrolled or uncontrollable. Lightning is the usual example, for it is both uncontrollable and unpredictable, for most practical purposes. The outcome of battle between evenly matched armies is often attributed to chance because the relevant factors are so multitudinous, many unrecognized, many out of control, that the issue is unpredictable, and some trivial matter (e. g., the horseshoe nail) may appear to tip the balance.

The influence of the aleatory on achievement may be studied on the campus with some success because achievement is measured in the classroom 1) frequently, 2) in standardized units, 3) under controlled and relatively uniform circumstances, and 4) is reviewed and recorded systematically. Such appraisal of chance factors is hence simplified by the academic machinery, as opposed to the largely intuitive and informal processes by which parents, employers, spouses, and diplomats usually estimate achievement and the factors associated with it. It is no easier here than elsewhere to measure the influence of crude undifferentiated luck, like breaking an arm or losing a textbook just before exams. But certain classes of individually unpredictable events can be treated profitably on a probability level, and three of these will now be considered.

Hard vs. Easy Instructors

The first of the three, designated Factor A, refers to the toughness or easiness of the instructors which the student draws in the semesterly grab bag of registration. A tough instructor, called "low", is one who tends to give his students low grades, in contrast to his easy or "high" colleagues. The data, for a recent fall semester, are taken from the records of a rather large institution of good reputation. Two hundred and six instructors of freshman-sophomore (lower division) classes and 179 instructors of junior-senior (upper division) classes, gave grades as shown in Tables I and II. Each value is the average grade, converted to numbers, given to a given instructor's lower (or upper) division students by him, counting A's as four, B's as three, C's as two, D's as one, and F's as zero grade-points.

TABLE I

GPA'S GIVEN BY 206 INSTRUCTORS OF LOWER DIVISION COURSES*

Lowest Instructor	1. 82 (High D)
Highest Instructor	3. 88 (High B)
Range	2. 06
Mean	2. 50 (Middle C)
17th percentile	2. 22
33rd percentile	2. 36
Median	2. 44
67th percentile	2. 55
83rd percentile	2. 82

*After excluding three instructors with fewer than 10 lower division students.

TABLE II

GPA'S GIVEN BY 179 INSTRUCTORS OF UPPER DIVISION COURSES*

Lowest Instructor	1. 97
Highest Instructor	3. 71
Range	1. 74
Mean	2. 79
17th percentile	2. 42
33rd percentile	2. 60
Median	2. 76
67th percentile	2. 96
83rd percentile	3. 20

*After excluding 33 instructors with fewer than 10 upper division students.

It is apparent that an entering naive freshman who gets X instructor rather than Y may thereby be raising (or lowering) his expected grade by as much as two letters. Inasmuch as the array is not normally distributed the standard deviation (though computed to be . 33) may not be used for estimating probabilities. If, however, the 206 instructors

are divided into the top, middle, and low thirds, the median GPA of the low third (2.22 at the 17th percentile) is a reasonable estimate of the most probable grade of the student in the unlucky low third. Similarly, the median GPA within the top third (2.82 at the 83rd percentile) is a reasonable estimate of the most probable grade of the student in the lucky top third.

There is therefore an equal probability, .33, that a student will fall into the middle group, neither advantaging nor disadvantaging him, perceptibly, or into the top group with a probable advantage of .38 of a letter (2.44-2.82), or into the low third, with a probable disadvantage of .22 of a letter (2.44-2.22).

The usual student takes five three-hour courses. What is now the probability of his gaining the 38/100ths of a letter or grade-point in each of his five courses? It is obviously 1/3 multiplied by itself five times: 1/243, or .004115. The probability of similarly losing in all five is also .004115. On the basis that advantage in one course is cancelled by disadvantage in another (which is not exactly true, for .38 advantage would outweigh .22 disadvantage), the probability of gaining net advantage to the amount shown is given in Table III.

TABLE III

PROBABILITIES OF WINNING FROM +5 TO -5 ADVANTAGES BASED ON THREE EQUALLY PROBABLE OUTCOMES SCORED +1, 0, AND -1, CF FIVE INDEPENDENT EVENTS

Net Advantage	Fraction	Probability
5	1/243	.0041
4	5/243	.0206
3	15/243	.0617
2	30/243	.1235
1	45/243	.1852
0	51/243	.2099
-1	45/243	.1852
-2	30/243	.1235
-3	15/243	.0617
-4	5/243	.0206
-5	1/243	.0041
TOTAL	243/243	1.0001

An informed guess can now be made about the importance of Factor A in distorting the student's grade-point average as a measure of his achievement. It may be said that out of 1,000 entering freshmen, about four or so who have earned, say, 35 grade-points on 15 units of classes, will actually be credited with an extra .38 points per unit in all their five classes, so raising their grade-points from 35 to 40.7. About 20 will similarly gain in four of their five classes, raising their points from 35 to 39.6. Some 62 will win 3.4 extra points, and 123 will win 2.3 points. The same numbers may expect corresponding disadvantages, but each disadvantage costs only .22 points per unit, so the total losses will be only three-fifths as great as the corresponding equi-probable gains. There remain some 580 freshmen whose net advantage is no more than plus or minus one, with grade-point of no more than 1.14 or loss of .66. [1*]

Before leaving Factor A it is noted that similar but less distortion may be seen in the differing grading practices among the 32 departments of the institution, where the interdepartmental range of lower division GPA's is 1.37 letters, and even among the nine divisions, where there is a range of .7 letters. In the same semester the proportion of A's to all lower division grades given by the various departments varied from 0.0 percent to 50.6 percent. These differences do not necessarily imply unfairness in grading. Perhaps the assumption of unfairness is justified prima facie, but many reasons might rebut it. Still, from the standpoint of the naive student, such variation represents chance, unpredictable and certainly uncontrollable. (Students do not remain naive long, but their sophistication is irrelevant here unless they are able to register for instructors they choose, something they can typically do only at the beginning of registration.)

The "Cutting Point" Error

Factor B is likewise universally present, seldom noted however, or systematically examined. This is the discontinuity distortion resulting from the fact that students' grades are recorded in a series of steps, either A or B or C or D or F, while their true performance if measured accurately would proceed on a continuum. The error introduced by reducing a continuum to a five point scale could be computed, but it is chosen to compare a 15 point scale with a five point scale, both because it is intuitively clear, and because a 15 point scale is actually practical. For the latter one need only add the pluses and minuses to the central-valued letters, recording B-minus and B-plus, as well as B, for example.

The distortion lies in the fact that in the process of classification all individuals are treated in a class as if, contrary to fact, they all possess the criterion characteristic to the same degree. For example, student X may get two B's and three C's, while Y may get three B's and two C's. If however X's are high B's and C's, and Y's are low B's and C's, X may be performing better than Y, as their GPA's would show if they were calculated with different numerical values for the pluses

*Footnotes will be found at end of article.

and minuses. In grade-points, the student who earns a C-plus and is credited with a C has lost one-third of a point per credit unit, while the C-minus student has gained an unearned one-third point. Inasmuch as classes usually carry three units of credit, the student loses or gains one even grade-point per course, with each plus or minus unrecorded.

The probability calculations parallel those in the previous pages. The probability that a student's semester performance belongs in the upper third of any letter range, that is , with a "plus" letter grade is 1/3; the probability that he is in the lower "minus" third is 1/3; and the probability that he is in the middle no-modifier range is 1/3. Reading from Table III, the probability that a student will be advantaged in all five of his classes by five minuses being unrecorded is .0041. The probability of four such advantages is .02, and of three is .06, remembering that lost minuses and pluses can cancel each other out, and the advantages counted are net.

Returning now to the 1,000 entering freshmen, it may be expected that as a result of Factor B, four of them earning, say, 35 grade-points in their 15 units of classes may actually be credited with five extra grade-points, some 40 in all. Twenty may win four, 62 or so may win three, and 123 may win two. The same numbers may lose the same number of points, leaving some 580 neither advantaged nor disadvantaged by more than one grade-point.

Lucky and Unlucky Guessers

Factor C, logically prior to B, has been relegated to last place because it is a somewhat more involved story. It has to do with the actual process of measuring achievement through the customary written examination. The imprecision of such measurement is notorious among teachers if not among students. This is especially true of the essay examination. A number of studies have concluded that an examiner would rank order his essays quite differently a second time from the first, and that a second examiner, equally familiar with the material, would also rank order the essays differently from his colleague, with probably a still greater degree of disparity. The disparity is different for different kinds of material, e.g., science as opposed to humanities, and for different instructors.

We have chosen for our purposes to review the objective type of examination where this source of error is presumed to be at its barest minimum. And indeed it is, for objective type tests enjoy practically perfect reliability. Clerks may safely be entrusted with scoring, for scoring is mechanical—hence the widespread adoption of such tests, not only in the classroom but also in civil service, industrial personnel work, and generally in psychometrics. But perfect scoring reliability distracts attention from another source of error which has not, so far as we know, been examined with any care. The writer is not referring to the problem of validity, as serious as that is, but to the errors in rank ordering which result from the operation of chance, as students guess at the answers which they do not know.

To make the point clear, an extreme example is chosen. Assume, for ease of calculation, a class of 10 students, taking an examination of 100 true-false questions. Rank-ordering and grading are to be done on the basis of the number of correct answers (right minus wrong makes no difference). Assume further that the average student knows 90 of the answers, and the range among the students is from 85 to 95 "knows." On the remaining questions they guess. To carry out this little exercise the writer "guessed" the proper number of times for each student, using a table of random numbers, with odd numbers "correct" and even numbers "incorrect." The results of carrying out this simulated test situation three times are shown in Table IV.

One will immediately notice the discrepancies between the true rank ordering by the students' knowledge, derived from column 2, and the new rank ordering resulting from the addition of chance results, columns 5, 7, and 9. The correlations between the values in column 2 and those in columns 5, 7, and 9 are respectively -.101, .509, and .816. R^2 values for the latter two are .259 and .665, indicating that some 26 percent and 66 percent of the variation in the final scores is accounted for by variation in the students' knowledge. The balance, 74 percent and 34 percent, is accounted for by chance. In the first run the discrepancy is of course most marked of all, with a slight negative correlation between the "know" score and the final score.

In this fictitious situation the small size of the population, the narrow range of "know" scores, and the fortunes of the random numbers table led to an uncommon degree of distortion. A more realistic estimate of distortion was sought by setting up a series of situations more like the everyday classroom on examination day. Twenty-five examination patterns were worked out and analyzed, as reported in Table V. Each situation was alike in that there were 25 students, each taking a 100 question true-false test, grades determined by number of correct answers less the numbers of errors. The 25 classes were divided into five groups, of five classes each, by the postulated numbers of "know" answers, with highs from 95 to 75, medians from 80 to 50, lows from 70 to 30, and ranges from 20 to 60. In classes one through five, for example, the best student knew 95, and guessed 5; the median student knew 80 and guessed 20; and the lowest student knew 50 and guessed 50. Column five gives the product

moment correlation coefficients from which the coefficient of determination may be computed. Although the data are surprisingly homoscedastic, rank difference coefficients also are given (column 6). The next four columns show the manner in which students' place in the rank order of the class are changed by their luck in guessing. Column 7 shows the mean change in rank, either up or down, and may be considered one measure of the error introduced by guessing. Column 8 shows the maximum change for any student in each class. Columns 9 and 10 show the mean absolute changes in rank for the top five and lowest five, respectively, of each class of 25 students; the upward movement by lucky poor students is about a third greater than the downward movement of unlucky good students.

Using as a measure of distortion the changes in rank order experienced by these 625 fictitious students in their 25 classes, such changes are found distributed as in Table VI. From this array of data it is seen that the two most favorably advantaged students moved upward in the rank order of their classes thirteen and one-half and 13 places, respectively. Forty-seven students moved upward by five or more places. The most seriously disadvantaged student fell 14 places, and 71 students fell five or more places. The curve describing this distribution of advantages and disadvantages is, though quite irregular, close enough to normal to permit calculation of probabilities from the table of the areas of the normal curve. The value of the standard deviation is 3.71. Generalizing, then, from this experience, the best estimate would be that 15.87 percent of a population of students would be advantaged by rising 3.71 places or more, and the same proportion would be similarly disadvantaged. One third, the luckiest third, would be advantaged by rising 1.60 or more places, and the unlucky third would be similarly disadvantaged by falling at least 1.60 places in the rank ordering of the class. To make the figures comparable to those found with Factors A and B, it may be estimated that the median gain or loss for those in the lucky or unlucky thirds, respectively, would be 3.59 places (corresponding to the values at the 17th and 83rd percentiles in the array of gains and losses).

The estimate of grade-point gain from gain in place-rank depends of course on grading practices. An ordinary distribution for lower division classes is 10 percent A's, 25 percent B's, 45 percent C's, 15 percent D's, and 5 percent F's, with a mean spread per letter of 20 percentage points. A rank order change of 3.59 places among a population of 25 corresponds to a change of 14.36 places in 100, which very roughly would be three-quarters of a letter, depending largely on the location of the change within the array.

We may hesitantly say then that the median advantage which the most fortunate third of a class will gain through Factor C is about 3/4 of a letter in any examination, with an equal loss for the least fortunate third. The impact of this on a student depends also on the number of examinations he has during a semester, for there will be an evening-out process going on, with usually some good luck compensating for bad, and vice versa (see the discussion below of this important point). If it is estimated that Factor C may raise or lower a student's semester grade by 1/3 of a letter, or one grade-point per three unit course, with respect to the top and bottom thirds of the class, this will be both a conservative estimate, and a most convenient one, inasmuch as it is identical in value with Factors A and B. In the discussion that follows one should, however, bear in mind that this may be an overstatement for C students, whose percentile scores cover such a wide range, and is probably an understatement for all others.

Turning for the third time to Table III, the probability is .004 that a student will enjoy this one grade-point gain, via Factor C, in each of five courses, .02 that he will in four of the five, and so forth.

<u>Joint Effects of Factors A, B, and C</u>

The time has now come to consolidate the data. What kinds of probability statements can be made of the joint effects of Factors A, B, and C?

If the probability of being in the most fortunate third in five classes with respect to any one factor is 1/243, or .0041, then the probability of being so fortunate for two factors simultaneously, in all five classes (gaining 10 advantage points) is (.0041) (.0041) or .000017, and the probability of such good fortune for all three factors, all five classes (with 15 advantage points) is $.0041^3$ or .00000007, more exactly, one out of 14,348,907. This lucky one in fourteen million students will have won but not earned, in one semester of five three-unit courses, five grade-points via Factor A, another five via Factor B, and a third via Factor C. If he is a middling C student, earning 35 points in 15 units of classes, he will actually rack up some 50 points on our records. Another equally unfortunate student will similarly lose.

These are indeed remote possibilities, hardly worth considering. Yet, for each student so heavily affected by the gift or loss of 15 grade-points, there will be 15 others winning or losing 14 (net) points, 120 who win or lose 13, and so forth.

Table VII gives the complete story toward which we have been moving. It looks as if perhaps about one out of eight of the students experiences no net gain or loss with respect to these three chance factors as defined (see column 3). Another quarter of them either gained or lost only a single point. Almost a third, 32 percent, may be expected to gain two or more points (see column 4), and another third to lose two or more. Likewise 13.5 percent may be expected to win at least four points, and another 13.5 percent to lose them.

TABLE IV

SCORES OF 10 STUDENTS ON THREE FICTITIOUS EXAMINATIONS AS AFFECTED BY LUCK IN GUESSING

1	2	3	4	5	6	7	8	9
Subj.	No. he knows	No. he guesses	No. guessed right 1st run	Total score run 1	No. guessed right 2nd run	Total score run 2	No. guessed right 3rd run	Total score run 3
A	85	15	12	97	9	94	9	94
B	87	13	7	94	7	94	3	90
C	89	11	6	95	4	93	5	94
D	90	10	7	97	3	93	5	95
E	90	10	7	97	5	95	6	96
F	90	10	5	95	8	98	3	93
G	91	9	3	94	5	96	4	95
H	91	9	4	95	7	98	4	95
I	92	8	3	95	6	98	6	98
J	95	5	1	96	1	96	2	97

TABLE V

CHARACTERISTICS OF 25 FICTITIOUS CLASSES, OF 25 STUDENTS EACH, SHOWING CHANGES IN INTRA-CLASS RANK ORDER RESULTING FROM GUESSING IN AN EXAMINATION

1	2	3	4	5	6	7	8	9	10
Class No.	"Know" Scores			Correl. Coef. r	Rank Dif. Coef.	Absolute Change Rank Order			
						Mean for Class	Max. in Class	Mean for	
	Max.	Med.	Min.					Top 5	Low 5
1	95	80	50	.915	.919	2.24	7.5	2.0	1.8
2	95	80	50	.866	.878	2.88	8.5	1.9	1.4
3	95	80	50	.888	.885	2.64	-8.0	1.6	2.6
4	95	80	50	.903	.887	2.56	-9.5	1.9	3.2
5	95	80	50	.907	.908	2.40	-6.5	2.1	1.8
6	90	80	70	.719	.696	4.72	13.5	3.2	4.6
7	90	80	70	.779	.844	3.28	8.5	3.9	4.3
8	90	80	70	.879	.886	2.28	-10.0	.5	1.4
9	90	80	70	.926	.882	2.68	-9.5	1.6	1.9
10	90	80	70	.803	.824	3.32	-9.5	3.3	3.2
11	90	60	30	.931	.922	2.12	-9.5	1.4	3.0
12	90	60	30	.839	.907	2.36	-8.0	1.5	2.2
13	90	60	30	.948	.944	2.04	5.0	.8	1.8
14	90	60	30	.960	.959	1.48	-6.5	.9	2.0
15	90	60	30	.924	.920	2.00	-9.5	1.1	3.2
16	75	60	40	.839	.862	2.76	7.5	2.8	2.6
17	75	60	40	.799	.790	3.84	-10.0	3.4	2.7
18	75	60	40	.798	.821	3.68	-9.0	3.2	2.1
19	75	60	40	.695	.740	4.28	13.0	3.0	3.0
20	75	60	40	.837	.800	3.12	-14.0	2.8	2.0
21	80	50	30	.917	.884	2.40	-9.0	.7	3.2
22	80	50	30	.913	.902	2.32	-8.0	1.0	2.8
23	80	50	30	.876	.799	3.48	-9.5	.4	5.1
24	80	50	30	.916	.915	2.24	-6.5	1.6	2.0
25	80	50	30	.888	.909	2.60	6.0	2.6	3.0

TABLE VI

CHANGES IN RANK ORDER OF 625 FICTITIOUS STUDENTS ORDERED INTO 25 ARRAYS ("CLASSES") OF 25 STUDENTS EACH RESULTING FROM CHANCE FACTORS IN GUESSING IN OBJECTIVE TYPE EXAMINATIONS

Amount of Change, in Places	Number of Students
+15, 15.5	0
+14, 14.5	0
+13, 13.5	2
+12, 12.5	0
+11, 11.5	0
+10, 10.5	0
+ 9, 9.5	1
+ 8, 8.5	8
+ 7, 7.5	12
+ 6, 6.5	11
+ 5, 5.5	13
+ 4.5	17
+ 4	24
+ 3.5	16
+ 3	22
+ 2.5	28
+ 2	33
+ 1.5	31
+ 1	53
+ .5	31
0.0	62
- .5	31
- 1	35
- 1.5	24
- 2	29
- 2.5	18
- 3	16
- 3.5	16
- 4	10
- 4.5	11
- 5, 5.5	23
- 6, 6.5	18
- 7, 7.5	8
- 8, 8.5	11
- 9, 9.5	8
-10, 10.5	2
-11, 11.5	0
-12, 12.5	0
-13, 13.5	0
-14, 14.5	1
-15, 15.5	0

TABLE VII

PROBABILITIES OF WINNING FROM +15 TO -15 NET ADVANTAGES BASED ON THREE EQUALLY PROBABLE OUTCOMES, SCORED +1, 0, AND -1 OF THREE INDEPENDENT SERIES OF FIVE INDEPENDENT EVENTS EACH

1	2	3	4
Net Advantage Points	Combinations of Events Giving Specified Points	Probability (Column Two Divided by 14,348,907)	Cumulative Probabilities
+15	1	.00000007	.00000007
+14	15	.00000105	.00000112
+13	120	.00000836	.00000948
+12	665	.00004634	.00005582
+11	2835	.00019758	.00025340
+10	9828	.00068493	.00093833
+ 9	28665	.00199771	.00293604
+ 8	71955	.00501467	.00795071
+ 7	157950	.01100781	.01895852
+ 6	306735	.02137689	.04033541
+ 5	531531	.03704331	.07737872
+ 4	827190	.05764829	.13502701
+ 3	1161615	.08095495	.21598196
+ 2	1477035	.10293711	.31891907
+ 1	1704510	.11879023	.43770930
0	1787607	.12458140	.56229070
- 1 to -15	6280650	.43770930	1.00000000
TOTAL	14348907	1.00000000	

It is of interest to note how inconsistent these figures are with the half thought-through notion that the more chance factors there are the more likely that the gains and losses will cancel, or at least shrink. The contrary is visibly true in these tables. The probability of being the most generously blessed person, lucky in all five classes, is in these examples 1/243 or .0041 for one factor, $.0041^2$ (or .000017) for two factors, and $.0041^3$ (or .00000007) for three factors. The probabilities of these substantially increasing maximal gains occurring do indeed shrink with the addition of more chance factors. But the probability of a given or constant gain occurring does not shrink; it increases. For example, the probability of gaining five advantage points as defined is, with respect to one factor, 1/243 or .0041 as stated. The probability of gaining five or more net advantage points through the operation of two factors is .040, and through three factors is .077.

It is important to distinguish here between an absolute number of advantages, i.e., grade-points, and a proportion, i.e., the grade-point average. As the semesters accumulate, chance will increase the absolute number of accidental grade-points lost or gained, but not in proportion to the total units attempted, on which the grade-point average is based. This does not, however, apply to increasing the number of chance factors operating in a given time period; for here the number of units of work is constant (not increasing with the passage of time) and hence any increase (or decrease) in grade-points, with the addition of more sources of chance, will also increase (or decrease) the grade-point average.

This is an exact parallel to flipping coins. In a million flips one gains or loses much more money than in 10, but the proportion of heads to flips approximates .500000 ever more closely: that is, the distorting effect of luck decreases, measured by that proportion.

Summary

Although the influence of fortuitous events generally may be impossible to estimate, even in the controlled situation of the classroom, at least an informed guess can be made as to the influence of certain classes of them. This has been done for three classes, (A) strictness of grading practices of the instructor, (B) loss of information inherent in the classifying process, and (C) guessing answers to objective type examination questions. The data do not permit calculation of the proportion of variance in grade-point average which may be so accounted for, but it is at least clear that the influence of these factors on recorded school achievement is far from trivial.[2]

FOOTNOTES

1 This is, of course, the crudest of approximations, subject to many unknown or unmeasured errors. It may be worth pointing out, however, that the process of dividing instructors into only three classes, i.e., the top, middle, and low thirds, serves to underestimate the influence of this factor. Six classes of toughness, for example, would both increase the job of calculating and increase the apparent distorting effect of this factor on GPA as a measure of achievement.

2 One may speculate on the influence of essentially these same error sources, i.e., kinds of luck, on non-academic achievement generally. This is done in the course of a more refined mathematical analysis of chance events in another as yet unpublished paper titled "A Probabilistic-Aleatory Approach to Achievement."

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

STUDENTS' USE OF SUMMER TIME AT THE UNIVERSITY OF WISCONSIN

L. J. LINS, C. A. SCHOENFELD,
ALLAN P. ABELL, F. CHANDLER YOUNG
The University of Wisconsin

WHAT DOES the average college student do during the summer? Does he go to summer school; travel; work; or simply take a vacation? If he works, how much does he save toward educational expenses? If he goes to summer school, what is the scope of the program he pursues? In short, what in fact are the present characteristics of the basic clientele available to a university as it seeks to offer optimum educational services throughout the year? What possible time and finances do students, enrolled during the regular year, have for summer attendance?

Such questions are fundamental to any discussion of academic calendar review; yet, surprisingly enough, they have not generally been asked in any systematic way. Educators sometimes have operated on hunches based upon partial data. Many hunches turn out to be incorrect when exposed to the light of actual statistics.

This is a report of what one university, The University of Wisconsin, Madison campus, discovered when it sought information on summer use of time on the part of its students in order to evaluate more effectively institutional policies and plans.* Every student registering on the Madison campus of The University of Wisconsin in the fall of 1961 was asked to complete a short punched card questionnaire. Some 94 percent of the students completed the cards. From the 18,847 usable cards, a variety of pertinent tables was constructed.

At the outset, let it be said these tables reaffirm a fact well known to education researchers; there is no such thing as a "typical student." Distinctions must be drawn on the basis of certain characteristics.

Summer Activities

The summer activities of University of Wisconsin students enrolled during the 1961-62 regular year centered around employment, summer school attendance, vacation at home or abroad, military camp, independent research, and a miscellany of other activities. The average student was engaged in a combination of activities. The way he divided his time differed markedly depending on a variety of factors.

By Class Level. There were notable differences between graduate and undergraduate students in their patterns of summer activities. More undergraduate students were employed during the summer than were graduate students — 85 per cent contrasted to 59 percent. The opposite was true for summer school attendance — undergraduate, 19 percent; graduate, 53 percent. The average undergraduate student was employed a week longer than was the average graduate student. Employment for the undergraduate and summer school for the graduate student did not, however, account for as much of the summer as might be expected. Only 60 percent of the undergraduate students were employed more than eight weeks, and only 44 percent of the graduate students had five or more weeks of summer school. Some 29 percent of the undergraduate and 34 percent of the graduate students had more than four weeks of vacation.

Undergraduate students showed increasing participation in summer school at each level from the freshman to the senior year: freshmen, seven percent; seniors, 33 percent. Continuing under-

* For the complete report see L. J. Lins, C. A. Schoenfeld, Allan P. Abell, F. Chandler Young, *Use of Time during 1961 Summer by Students Registered during 1961 Fall Semester, The University of Wisconsin, Madison Campus* (Madison: The University of Wisconsin, Office of Institutional Studies, May, 1962).

graduate students were more apt to attend summer school than were new advanced standing students or new freshmen.

By Sex. Undergraduate men were more apt to work than were undergraduate women: 91 percent versus 77 percent; there was a greater decrease in the proportion with summer employment from the freshman through the senior years for women than for men. A slightly higher proportion of graduate men than graduate women was employed: 61 percent as compared with 51 percent. Graduate and undergraduate men with employment worked about one week longer on the average during the summer than did the women.

Undergraduate women and graduate men were somewhat more apt to attend summer school than were the men or women in the other respective classifications. Among professional students, a higher proportion of men than women attended summer school.

Within each classification, men attending summer school had less vacation time while women attending summer school had less weeks of employment.

By Marital Status. The married undergraduate student was more apt to attend summer school than his single counterpart, less apt to take a vacation, and somewhat less apt to be employed. A similar pattern was exhibited by married graduate and professional students.

By College or School. Undergraduate summer employment ranged from 96 percent in the College of Agriculture to 64 percent in the School of Nursing. Undergraduate summer school attendance ranged from 58 percent in Nursing to 10 percent in Agriculture.

Commerce had the largest percentage of graduate students with employment (68%) and the lowest percentage with summer school attendance (34%). The School of Pharmacy had the highest percentage in summer school (74%) and the lowest percentage with employment (44%). Graduate students in Agriculture had a relatively high participation in both summer school (73%) and employment (60%).

By Home Address. Students from Madison and from out-of-state were somewhat more inclined to attend summer school than were students from other parts of the State of Wisconsin. Students from Wisconsin were more apt to be employed during the summer than were non-Wisconsin students. This may be due to varied needs for summer earnings in order to continue in the fall.

By Cumulative Grade-Point Average. A slightly greater proportion of undergraduate students with low cumulative grade-point averages attended summer school than of students with high cumulative GPA's. The size of the difference does not appear great enough, however, to weaken the quality of the summer session student body.

Summary. There are 14 weeks between the close of the spring semester at Madison and the opening of the fall term. Of these 14 weeks, the composite undergraduate male spent nine and one-half weeks in gainful employment, two and one-half weeks on vacation, one week in summer school, and the balance in a variety of other activities. His coed counterpart worked for seven and one-half weeks, vacationed for four and one-half weeks, spent one and one-half weeks in school, and divided the balance of time among miscellaneous activities. The composite graduate student exhibited a different pattern: male, five and one-half weeks of employment, four weeks of summer school, and three weeks of vacation; female, four and one-half weeks of vacation, four weeks of employment, and three and one-half weeks of summer school.

Looked at another way, 91 percent of the UW undergraduate men, 77 percent of the undergraduate women, 61 percent of the graduate men, and 51 percent of the graduate women reported being engaged for a time in some type of summer employment. Summer school attendance involved 16 percent of the undergraduate men, 24 percent of the undergraduate women, 48 percent of the graduate men, and 54 percent of the graduate women.

The majority of the undergraduate students (80%) had some combination of summer activity that included employment and vacation while only 19 percent had a combination that included summer school. In contrast, 53 percent of the graduate students included summer school in their summer activities.

Summer School Attendance

As has been cited, summer school attendance was related to classification: undergraduate students, 19 percent; professional students, 25 percent; graduate students, 53 percent. As might be expected, the majority of UW students enrolled during the regular year and who took summer course work did so on the Madison campus: undergraduate, 12 percent; professional, 22 percent; graduate, 47 percent. Other institutions attracted measurable numbers of UW students: undergraduate, five percent; professional, two percent; graduate, six percent. Only a very small number of Madison campus students pursued summer work either at the University of Wisconsin-Milwaukee, through correspondence study, or at one of the University Centers.

There was a general trend from the sophomore to the senior level of decreasing summer school

attendance on campuses other than The University of Wisconsin at Madison, with an increase from the freshman to the senior level in UW correspondence study, and a much greater increase in proportions attending the Madison summer sessions. These trends were more pronounced among women than among men.

Credits Earned by Class Level. Among the students who earned academic credits during the summer, the mean varied from 6.6 for professional, to 5.6 for undergraduate, to 4.1 for graduate students. Among undergraduate students, freshmen had the lowest mean and seniors the highest.

By Sex and Marital Status. Women tended to carry more credits than did men in all classifications with the exception of law and medicine. Of the 1,548 graduate men in the questionnaire group who attended the Madison summer sessions, less than half earned more than three credits.

The influence of marital status upon total credits earned was insignificant.

By College or School. Total earned credit means and medians by college or school for undergraduate students were fairly uniform. Graduate student medians varied from six credits in Education and Home Economics to two credits in Agriculture.

By Home Address and Cumulative Grade-Point Average. Only minor differences were exhibited between Wisconsin and non-Wisconsin students in total earned credit means and medians. Cumulative grade-point averages likewise appear to have had no significant effect on the number of credits carried.

Summary. The average UW undergraduate male registered at Madison during the first semester earned .88 academic credits in summer; the undergraduate female, 1.32; the graduate male, 2.09; and the graduate female, 2.28.

Among those students reporting summer school attendance, the typical undergraduate man and woman each earned six credits; the graduate woman, five credits; and the graduate male, three credits.

Summer Employment and Savings

Employment accounted for the greater portion of the summer activity of the majority of University of Wisconsin students, but there were wide variations in amounts of savings.

By Class Level. Among the undergraduate students, 86 percent were employed, but 14 percent had no savings, leaving 72 percent with an economic profit for the summer. Ninety percent of the savings were within a range of $97 to $919. The median savings for the undergraduate students employed was $379; the mean was $419. The average savings for all undergraduate students was $300. The freshman group had the lowest mean savings, the junior group the largest.

Among graduate students, 60 percent were employed, but 29 percent had no savings, leaving 31 percent with summer savings. The range of 90 percent of the savings was from $101 to $1,027. The median savings for graduate students employed was $352; the mean was $428. The average savings for all graduate students was $132.

Among professional students, 86 percent were employed; 24 percent had no savings, leaving 62 percent with summer savings. Ninety percent of their savings were within a range of $145 to $1,200. The median savings for those employed was $515; the mean was $548. The average for all professional students was $339.

By Sex and Marital Status. Mean savings were higher for the men than for the women in all three classifications. A higher proportion of single students than of married students had savings; however, the savings mean of single students was less.

By College or School. The undergraduate students in Engineering and Agriculture had the highest mean savings. Commerce undergraduate students had the highest median. The Engineering group also had the highest mean savings among graduate students. These differences may be more a function of sex than of college or school in which enrolled.

By Home Address. The smallest percentage with summer savings was the group of undergraduate students from regions other than Wisconsin. The graduate student group from Wisconsin areas other than Madison had the highest mean savings of any group. Differences in savings probably were due largely to the varying needs for money to continue education during the regular year.

By Cumulative Grade-Point Average. The lower a student's cumulative grade-point average, the more apt he was to be employed in the summer and and the less apt he was to have savings. The differences between students with high CGPA's and those with low CPGA's were not marked, however.

Summary. There was a wide variation by class in the typical summer savings of the students with savings and enrolled at Madison during the first semester 1961-62, with the savings being $308 for graduate females, $406 for graduate males, $294 for undergraduate females, and $437 for undergraduate males.

Summer Savings Reserved for Education

The majority of University of Wisconsin students with summer savings applied most of these savings toward educational expenses.

By Classification. The undergraduate group had the greatest percentage of persons who reported savings but who did not intend to use these savings for education (7%). However, the undergraduate group still had the highest percent (65%) with applied savings; the undergraduate mean was $396. The graduate group with savings exhibited the greatest difference between mean savings ($428) and applied savings ($391), while professional students had the largest mean applied savings ($513).

By Sex and Marital Status. A smaller percent of the women than of the men had applied summer savings. Married students were able to save less for educational expenses than were single students.

By College or School. The pattern of proportions with applied savings by college division closely paralleled that of the proportions with savings. Undergraduate students in the School of Home Economics exhibited the greatest disparity between savings and applied savings; in the graduate student category the difference was most marked among students in Education.

By Home Address. A higher proportion of the undergraduate students from regions other than Wisconsin had savings not applied to educational expenses than of undergraduate students from Wisconsin. The Other Wisconsin graduate group, like the Other Wisconsin undergraduate group, had the greatest percentage of individuals with applied savings.

By Cumulative Grade-Point Average. The higher a student's cumulative grade-point average the less apt he was to save toward educational expenses. The medians ranged from $353 for undergraduate students with 3.00 or over to $400 for undergraduate students with 2.29 or under. A higher proportion of the high cumulative grade-point average group than of the low cumulative grade-point average group, however, had some savings.

Summary. The amounts applied toward educational expenses by the "average" student were as follows: graduate women, $81; undergraduate women, $161; graduate men, $118; and undergraduate men, $317. Among students reporting summer employment, the median savings toward educational expenses were: undergraduate men, $414; graduate men, $335; graduate women, $285; and undergraduate women, $265.

Generalizations

The data reported here can be interpreted in several ways. It was to answer a specific question, however, that the study was undertaken. These generalizations will confine themselves to that topic, namely, how many University of Wisconsin students might be available for year-around study?

Undergraduate Students. This study reaffirms the findings of other studies that, for a significant number of University of Wisconsin undergraduate students, summer employment savings may mean the difference between continuing or not continuing in school. Some 65 percent of the undergraduate students had applied savings, and half of these had savings of approximately $346 or more; this means that for at least 4,500 undergraduate students (32%) summer school attendance would have represented a serious loss in income.

On the other hand, this study reveals a significant number of UW undergraduate students for whom summer study would seem to have been economically feasible given modest scholarships and/or loans. There were nearly 4,000 undergraduate students with no summer savings and almost 5,000 with applied savings of less than $400 each. Of the students surveyed, only 2,330 spent more than four weeks in summer school, leaving a net potential of over 7,000 undergraduate students (52%) for whom eight weeks or more of summer school might have been feasible, given suitable curricula and motivation.

These summer school potentials were not evenly distributed by class, however. In general, there was a steady decline in their numbers from freshmen through seniors. Nor were they evenly distributed by sex and home address. Women earned less credits and money than men; nonresident students earned less credits and money than resident students did.

Graduate Students. While graduate students now make use of The University of Wisconsin, Madison campus, summer sessions to a much greater degree than undergraduate students do, here, too, there would seem to be an untapped potential. Over 70 percent of the graduate students reported no applied summer savings; an additional 16 percent had applied savings of under $400. Subtracting those who now spend over four weeks in summer school leaves a balance of 1,766 (43%) for whom eight weeks or more of summer school might have been possible. It should be kept in mind, however, that lack of summer savings of graduate students not attending the summer sessions may have been due to factors such as debt retirement, degree research away from the cam-

pus, and family responsibilities. A high proportion were married. Many had assistantships during the academic year but not during the summer.

For the 12 percent of the graduate students who had applied savings of \$400 or more, summer school attendance would seem to present a serious economic obstacle.

Professional Students. In general, professional students followed the undergraduate student pattern. Their numbers, in any event, were not significant.

Summary. From this study it would seem that there may be quite a number of University of Wisconsin students not now making optimum use of the 14 weeks of summer time in terms of either money or credits accumulated. Whether this potential can in fact be translated into summer attendance on the Madison campus will probably hinge on a variety of socio-economic and personal factors which educational institutions have only begun to explore. On the basis of present knowledge, it is not safe to assume that larger numbers of students would or could attend a summer session of any particular length or that establishing a summer session longer than any of the present sessions would automatically attract a satisfactory enrollment. It appears quite certain that to increase enrollments substantially during the summer it would be necessary to expand scholarship and/or loan programs and to explore other means of developing greater student utilization of summer offerings.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 2, December 1962)

STUDENT REACTIONS TO UNIVERSITY OF WISCONSIN SUMMER SESSIONS

L. J. LINS, C. A. SCHOENFELD,
ROBERT A. REES, ALLAN P. ABELL
The University of Wisconsin

PERHAPS NO ASPECT of American higher education is presently under more scrutiny than the college and university summer session.[1*] Perennially the topic both of "enthusiastic speeches"[2] and of "a great many suspicions,"[3] the summer session today is very much on the minds of those seeking optimum utilization of campus facilities around the calendar in the face of rapidly increasing enrollments.

The Study Background

As one of the institutions seeking to discover ways of maximum use of the summer as an integral part of year-around operation, the University of Wisconsin undertook in the summer of 1961 a depth analysis of its Madison campus summer students—their backgrounds, their aspirations, their needs, and their assessments. A more complete understanding of present summer session functions should aid in formulating future summer session policies.

While what emerges from any self-study of Wisconsin summer student posture and attitude may reflect geographic and psychological bias, this report should contribute to the literature on summer session problems and possibilities. At a minimum it identifies areas meriting further research; hopefully, it also identifies guidelines for future administrative action.

Study Framework

The study itself consisted of controlled interviews with 184 students selected at random from approximately 9,000 students enrolled in the 1961 Summer Sessions at Madison. On the basis of known factors, the sample was judged to be sufficiently representative of the population as a whole to warrant making generalizations based on the survey data. (A similar survey instrument was employed at the University of Minnesota in the summer of 1956.[4] The two studies offer interesting points of substantiation and divergence.)

The questionnaire employed sought answers to four broad questions:

1) Who is the summer student? What are his personal, educational, and vocational characteristics? Particularly, in what ways is he different from the regular-year student?

2) What is the place of summer session work in the student's total educational program? Why do students attend the Summer Sessions and through what sources do they secure information about the summer programs? How does summer session work contribute toward degree progress?

3) How do students appraise their summer session experiences? What do they think about their courses? How do they judge them in comparison with those they have taken during the academic year? What are their reactions to other phases of the summer session program? What do the students believe to be particularly outstanding or particularly unsatisfactory about the Summer Sessions?

4) What length of session has the greatest utility? Is a four-week session too short; a summer semester too long? Does the eight-week session meet student needs? What opening and closing dates are most convenient?

In the following summary, percentages are rounded to the nearest integer. For more precise figures and a more complete treatment of the findings, the reader is referred to the complete report.[5]

The Summer Student

The "typical" summer student bears little resemblance to the conventional image of "Joe College." He is quite apt to be married. He is over 24 years of age. He is working toward an advanced degree. He either went to school or taught school the year before. He has attended summer school before or plans to go again. He may work besides attending classes. He prefers "the reading room" to other Wisconsin Student Union activities.

Classification

There are striking differences between the summer and winter student bodies at Madison on the

*Footnotes will be found at end of article.

basis of class level.

During the regular year, students working toward graduate degrees constituted about 23 percent of the total students enrolled. In summer the Graduate School percentage doubled— to 46 percent. (As a matter of fact, if the special students taking advanced work are included, over half of the summer students are at the graduate level. There were about as many graduate students enrolled on the Madison campus in summer as in winter despite the differences in total enrollment.)

Undergraduate degree candidates at the University constituted 72 percent of the winter population but only 28 percent of the summer. There were very few non-degree candidates during the regular year while during the summer over 24 percent of the students were not working toward a UW degree.

In terms of present functions, the heavy graduate enrollment in the summer dictates a curriculum rich in opportunities for independent investigative work in library and laboratory. In terms of future policies, the best chance for summer enrollment expansion may lie in encouraging an increasing number of undergraduates to continue on campus the year-around.

Sex and Marital Status

There was a slightly greater percentage of women in the summer student population than in the winter–38 percent and 34 percent respectively. The proportion of married females increased from three percent in winter to eight percent in summer; the percentage of single females was the same–30 percent. The percentage of married males increased from 16 percent in winter to 31 percent in summer. The percentage of single males dropped from winter to summer–from 50 percent to 31 percent.

These data are particularly relevant to persons responsible for housing the summer student body. Since a high proportion of single males work during the summer to provide some of the funds for regular year enrollment, 6 it probably will be necessary to supply scholarship and loan funds to replace summer job incomes if more single males are to attend the Summer Sessions.

Home Address

The percentage of students from Dane County was not markedly different in winter and in summer. The percentage of students from Wisconsin counties other than Dane decreased markedly from the fall semester to the Summer Sessions (47% and 38%). The rise in percentage of out-of-state students, from 30 percent to 37 percent, is accounted for largely by the fact that the University attracts many students from other institutions for the summer only.

Age

The summer student was strikingly older than his winter counterpart. In total, 49 percent of the summer student body was 25 years of age or older, contrasted to 19 percent in the regular year. The summer undergraduate was older than the winter undergraduate; in summer 72 percent were over 19, in winter 48 percent were. The summer graduate student was older than the winter graduate student; in summer 71 percent were over 24 years of age, in winter 61 percent were.

These statistics are of special significance to those planning summer co-curricular activities.

Occupation

What were the students doing during the year immediately prior to their summer attendance? Overall, 59 percent attended the University of Wisconsin, 11 percent attended another institution, 23 percent were teaching and/or were counselors or administrators, and four percent were employed in a non-school occupation. Each student classification departed from this average, however. The undergraduate student body was made up predominantly of continuing University of Wisconsin students (84%). The graduate students were either continuing University of Wisconsin students (61%) or were teachers and administrators from high schools and colleges (35%). Nearly half of the Summer Session Specials were students from other colleges.

Degree Work

As of July 1, 1961, almost two-thirds of the total sample held some degree. In fact, 40 percent of the graduate students had a master's degree. Of the total students in the sample, 86 percent were working toward a degree. Of the graduate students, 42 percent were working toward a doctorate. Thirteen percent of the sample planned to earn a degree during the 1961 Summer Sessions. Only two percent of the regularly enrolled UW undergraduate and graduate students expected to earn a degree at a college or university other than Wisconsin; 47 percent of the specials did.

Summer Session Experience

A majority of the graduate students had attended a previous summer session at Madison; a majority of the undergraduates and specials had not. Of the total group, 34 percent had attended a summer session at another institution; 15 percent of these had been in attendance on three or more campuses. Many of the persons attending the 1961 Summer Sessions expected to attend a future University of Wisconsin summer session.

Summer Employment

Of the students in the sample, over one-third had jobs in addition to attending classes: 45 percent of the undergraduate, 31 percent of the graduate, and 31 percent of the special students. Of the students with jobs, 65 percent earned less than \$400, and 17 percent earned \$600 or more. Graduate students earned more than undergraduate or special students.

Whether or not a student worked and how much he earned had little or no correlation with the number of courses he carried for credit.

Extra-Curricular Activities

Outdoor sports were popular with 71 percent of the students, with swimming being the most popular followed by tennis, golf, sailing and fishing. A like percentage of students was interested in plays, concerts, ballets, films, lectures, and art exhibits. Forty-four percent of the students expressed an interest in "Union Activities," with the largest number of votes going to "the reading room." All types of extra-curricular activities were more popular with undergraduate than with graduate and special students.

Summer Student Goals

Students obviously come to the Summer Sessions for a variety of reasons. While a psychologist could speculate that an individual might not either recognize or be willing to verbalize his real reason, it is nonetheless pertinent to ask him.

Reasons for Coming

It was found that nearly all summer students (90%) cited a "most important" reason related to an educational purpose: 39 percent, to accelerate educational progress; 17 percent, to take courses as a part of a balanced year-around program; 11 percent, to complete a program of studies largely through summer attendance; and 12 percent, to take work with a particular professor or to use the library or laboratory facilities.

Some 60 percent also gave a "most important" professional reason for attendance, the two most common being "to obtain a graduate degree" and "to enrich my professional background by further studies."

Only a third of the students listed a "most important" personal reason for attendance, the three most common being "to take courses purely for personal interest," "to be able to live at home while attending summer sessions," and "to meet new people."

Sources of Information

How do students obtain information about the Summer Sessions? Taken as a group, three informal means of communication account for 45 percent of the responses—talks with UW staff members, talks with students who have attended, and previous personal experience. The formal means of communication, accounting for 42 percent of the dissemination, is a combination of reading the Summer Sessions Bulletin and writing to the Summer Sessions Office for information. These data suggest that The University of Wisconsin summer program is its own best advertisement when supported by adequate literature.

Student Assessments

In the literature of higher education, summer instruction has at times been unkindly dealt with.[7] Despite growing evidence to the contrary,[8] there remains the lingering suspicion in some academic quarters that the summer session is a vermiform appendage where second-rate professors offer second-rate courses to second-rate students.

What do the students themselves think about this? It has already been pointed out that they are characterized by maturity and intellectual drive. Are they not in a valid position to contribute to Summer Sessions assessments? Many educators think they are.[9]

General Evaluation

The students were asked to estimate the value of attending the Summer Sessions in terms of their chief reasons for attending. They were given five choices: very well satisfied, fairly well satisfied, uncertain of the value, somewhat dissatisfied, and very dissatisfied. Of the students who replied to the question, 94 percent said they were either fairly well or very well satisfied. None of the students expressed complete dissatisfaction. A somewhat higher proportion of UW degree candidates expressed complete satisfaction than did special students.

Course Assessments

One of the questions asked was, "How much are you learning?" In every University college or school 76 percent or more of the students indicated that they were learning as much or more in their summer courses as they had learned in other courses taken previously. Some 40 percent of the students said they were learning more in summer. The responses of the UW degree candidates generally were more enthusiastic than were those of the special students.

The students were also asked, "How interesting are your courses?" Over 80 percent of the students felt that the Summer Sessions courses were either above average, more interesting than most, or among the most interesting classes they had ever

taken. There were no great differences between UW degree candidates and special students.

A third question asked was, "How would you rate the teaching procedures?" Again, over 80 percent of the students rated summer teaching average or better, and 23 percent said the instruction was "a model of good teaching." UW undergraduates tended to be slightly more complimentary than UW graduate students.

In response to the question, "Would you recommend this course to a close friend whose interests are like yours?" only 13 percent said they would not.

Comparisons

The continuing students in the sample were asked to compare summer courses with regular year courses at Madison. In terms of work requirements, nearly half said there was no difference, over one-fourth felt summer courses were more difficult, and over one-fourth said summer courses were less demanding. In terms of quality of instruction, three-fourths felt that professor performance was about the same, 18 percent thought it was better in the summer, and seven percent said it was generally poorer in the summer. In terms of amount learned, over half said they had learned about the same, less than one-fourth said they had learned more, and one-fourth said they had learned less. In general, UW undergraduates tended to rate summer courses higher in comparison with regular year courses than did UW graduate students.

Notable Satisfactions

Each student was asked if there was any phase of the summer program that he found particularly outstanding. The majority of students did not respond to this question. Among those who did, the most frequently cited factors were "diversified offerings," "general atmosphere and organization," "scheduling of classes," and "cultural and social acitivities."

Notable Dissatisfactions

Each student also was asked if there was any phase of the summer program that he found particularly unsatisfactory. Here again the majority of students did not respond. Among those who did, the most frequently cited factors were "concentrated instruction," "crowded schedule," and "unsatisfactory teaching."

It is interesting to note that some of the elements that were listed by some students as unsatisfactory, were listed by others as outstanding. Some of the responses were picayune, others thoughtful. Taken together, they do not seem to indicate any overwhelming agreement on either highlights or problem areas.

Missing Courses

The fact that the summer session administrator is dealing with a host of individual needs and slants became strikingly apparent when the students were asked to list any courses they would like to have taken but which were not offered in the Summer Sessions. Out of 64 responses, 61 different courses were suggested, ranging from advanced electrical engineering to Swahili. The only duplications were in the areas of botany, home economics, and history of education. The requests document both the need for a rich and varied curriculum and the difficulty in maintaining one that satisfies everyone.

Summer Session Patterns

"How long should a summer session be?" is a question being discussed frequently by educational administrators. Students have their own ideas.

Four-Week

Nearly half of the students consulted expressed an opinion on the length of the four-week session. Of those who did, only eight percent said it was "just right for my needs." Another 32 percent said it was "adequate." The remaining 60 percent felt that a four-week course was too short or too concentrated. The majority recommended that the four-week session start in June.

Eight-Week

Over two-thirds of the students indicated that a half-semester session was "just right" for their needs. However, 16 percent of the graduate students said it was too short. Over half of the students were satisfied with the present starting date although one-fourth would prefer to have it start earlier.

Twelve-Week

Over a third of the students in the sample said they would attend a 12-week session because it would allow them to carry more credits; 10 percent said "maybe." Nearly half of the students felt 12 weeks would be "too long or inconvenient." Sixty percent of the students felt a 12-week session should start as early as possible in June. Significantly, more UW degree candidates were in favor of a 12-week session than were special students.

Summer Semester

Slightly less than a third of the students in the sample said they would attend a summer semester,

and another eight percent said "maybe." Half of the students said "no." A higher proportion of UW degree candidates expressed an interest in a summer semester than did special students.

Were the same questions to be put to a group of students not in attendance in summer, the reactions might be quite different.

Interpolating from the data at hand, it is expected that many students would find acceptable a summer session pattern that offered a longer term in the form of two consecutive short terms.

Generalizations

Any study based on a two percent sample has its limitations, as does any study based on the subjective judgments of the interviewees. Granting these pitfalls, the study here reported, however, would seem to offer certain clues to administrative decisions and to further research about the Summer Sessions.

The growing impression that summer session students are more mature than those attending during the academic year is supported by the findings reported in this paper. The summer student is older. He is more apt to be married. He is likely a candidate for an advanced degree. He has sharply defined educational goals. His high academic motivation makes him less susceptible to campus diversions.

From the perspective of the student body enrolled, the Summer Sessions serve first of all as an integral part of a year-around program for students in continuous attendance. Secondly, particularly for Wisconsin teachers and school administrators, the Summer Sessions offer the only chance for periodic professional training. Thirdly, a more special function is the provision of an opportunity for college students and teachers from around the country to become acquainted with Wisconsin skills and resources.

The reactions of Summer Session students to the total program and to individual courses provide considerable evidence that the Summer Session is a worthy expression of university enterprise in maintaining academic standards and meeting student needs.

Current student opinions on a longer summer session indicate that while there is a measurable "built-in" clientele for a summer semester, a decided change of attitude will have to take place in a majority of students before summer session enrollments would approach those in the regular year. For a significant number of students now in attendance, only a half-semester would fit their schedules.

At least five research areas suggest themselves:

1) With respect to length, is there in fact an optimum instructional period?

2) What are the economic, social, and psychological barriers to increased summer attendance?

3) With respect to academic qualifications, what types of students typically avail themselves of summer offerings?

4) What are the past, present, and potential contributions of the summer session to the academic careers of degree candidates?

5) Are there opportunities for fruitful inter-institutional cooperation in curricular offerings?

At least two broad administrative guidelines emerge: 1) a consciousness of the Summer Sessions as an integral part of the year-around operations of the institution, and 2) a sympathy for the unique role of the Summer Sessions in meeting special needs.

FOOTNOTES

1 Clarence A. Schoenfeld, "The Summer Session Comes of Age," The Shape of Summer Sessions to Come (The University of Wisconsin: Summer Sessions Office, 1961), p. 1.

2 John E. Stecklein, Mary Corcoran, and E. W. Ziebarth, The Summer Session: Its Role in the University of Minnesota Program (University of Minnesota: Bureau of Institutional Research, c. 1958), p. 16.

3 Thomas E. Crooks, The Summer School in Higher Education (New York: Columbia University, 1961), p. 4.

4 Stecklein, Corcoran, and Ziebarth, op. cit.

5 L. J. Lins, C. A. Schoenfeld, Robert A. Rees, Allan P. Abell, Student Reactions to 1961 Summer Sessions, The University of Wisconsin, Madison Campus (Madison: The University of Wisconsin, Office of Institutional Studies, April 1962).

6 L. J. Lins, Student Expenses and Sources of Income 1960-61 Academic Year, The University of Wisconsin, Madison Campus (Madison: The University of Wisconsin, Office of Institutional Studies, October 1961).

7 Charles S. Slichter, "'Debunking' the Master's Degree," in Association of American Universities Journal of Proceedings and Addresses of the 29th Annual Conference, University of Chicago Press, 1927, pages 107-11.

8 Stecklein, Corcoran, and Ziebarth, op. cit.

9 Martha B. Lucas and I. D. Weeks, quoted in Clarence A. Schoenfeld, "The University and Its Publics," Harpers, 1954, p. 20 and p. 45.

CONTRIBUTORS

ALLAN P. ABELL is a Project Assistant, Office of Institutional Studies, The University of Wisconsin, Madison 6, Wisconsin.

O. F. ANDERHALTER is Director of the Bureau of Institutional Research, St. Louis University, St. Louis, Missouri.

GRANT A. BARNES is an Administrative Analyst for Analytical Studies, Office of the Vice President – Finance, University of California, Berkeley 4, California.

SAMUEL BASKIN is Director of Educational Research and Professor of Guidance and Psychology, Antioch College, Yellow Springs, Ohio.

PAUL S. BURNHAM is Director of the Office of Educational Research, Yale University, New Haven, Connecticut.

CLIFFORD L. CONSTANCE is Registrar, University of Oregon, Eugene, Oregon.

FRANCIS J. DONOHUE is Executive Director of the Self-Study, Fordham University, New York 58, New York.

JAMES DRASGOW is Lecturer in Psychology, University of Buffalo, Buffalo 14, New York.

EUGENE S. EDGINGTON is Associate Professor of Psychology, Kansas State Teachers College, Emporia, Kansas.

ROSA ESTHER ESCALERA is Research Officer and Instructor of Mathematics, The University of Puerto Rico, Rio Piedras, Puerto Rico.

E. B. FRED is President Emeritus of The University of Wisconsin, Madison 6, Wisconsin.

ROGER GRATWICK is Dean of Men and Chairman of the Medical School Appraisal Committee, University of Buffalo, Buffalo 14, New York.

D. J. GUZZETTA is Vice President and Dean of Administration, The University of Akron, Akron 4, Ohio.

IRMA HALFTER is Director of Psychological Testing, DePaul University, Chicago 4, Illinois.

CHARLES E. HOWELL is Director of Research of the Bureau of University Research, Northern Illinois University, DeKalb, Illinois.

ROBERT E. HUBBARD is Director of the Office of Administrative and Curricular Research, Wayne State University, Detroit 2, Michigan.

BERNARD C. KIRBY is Associate Professor, Department of Sociology and Anthropology, San Diego State College, San Diego 15, California.

RALPH E. KIRKMAN is Academic Dean, Mobile College, Mobile, Alabama.

ANN K. KNUDSEN is a graduate student at Fairleigh Dickinson University, Teaneck, New Jersey.

ALLAN A. KUUSISTO is Dean of the Graduate School, University of New Hampshire, Durham, New Hampshire.

GERHARD LANG is Assistant Professor of Psychology, Fairleigh Dickinson University, Teaneck, New Jersey.

L. J. LINS is Professor and Coordinator of Institutional Studies, The University of Wisconsin, Madison 6, Wisconsin.

ADELE M. MILLER is Assistant to the Dean, Brooklyn College of The City University of New York, Brooklyn 10, New York.

JAMES R. MONTGOMERY is Director of Institutional Research, University of Tennessee, Knoxville, Tennessee.

ROBERT R. RAMSEY, JR. is Director of the Office of Admissions and Freshman Scholarships, Yale University, New Haven, Connecticut.

ROBERT A. REES is a Woodrow Wilson Fellow and Teaching Assistant, The University of Wisconsin, Madison 6, Wisconsin.

SIDNEY G. ROTH is Director of the Office of Research Services, New York University, New York 3, New York.

JOHN DALE RUSSELL is Director Emeritus of the Office of Institutional Research, New York University, New York 3, New York (Current address is RR No. 10, Russell Road, Bloomington, Indiana).

C. A. SCHOENFELD is Associate Director of the Summer Sessions and Associate Professor of Journalism, The University of Wisconsin, Madison 6, Wisconsin

FREDERICK E. SCHWEHR is Physical Facilities Specialist, Board of Regents of Wisconsin State Colleges, State Capitol, Madison, Wisconsin.

AMEDEO G. SFERRA is Professor of Psychology and Chairman of the Department of Psychology, Fairleigh Dickinson University, Teaneck, New Jersey.

BRUCE STOCKIN is Instructor in Psychology, University of Buffalo, Buffalo 14, New York.

EDWARD M. STOUT is Registrar, De Paul University, Chicago 4, Illinois.

JOSEPH TANENHAUS is Assistant Director for Institutional Research, Office of Research Services, New York University, New York 3, New York.

LORING M. THOMPSON is Director of the Office of University Planning, Northeastern University, Boston 15, Massachusetts.

D. GORDON TYNDALL is Director of Analytical Studies, Office of the Vice President— Finance, University of California, Berkeley 4, California.

PHILIP H. TYRRELL is Director of the Office of Institutional Research, Rensselaer Polytechnic Institute, Troy, New York.

GERALD H. WHITLOCK is Professor of Industrial Management, University of Tennessee, Knoxville, Tennessee.

F. CHANDLER YOUNG is Assistant Dean of the College of Letters and Science, The University of Wisconsin, Madison 6, Wisconsin.

Publishers: File two copies of this Form with your postmaster.

STATEMENT REQUIRED BY THE ACT OF AUGUST 24, 1912, AS AMENDED BY THE ACTS OF MARCH 3, 1933, JULY 2, 1946 AND JUNE 11, 1960 (74 STAT. 208) SHOWING THE OWNERSHIP, MANAGEMENT, AND CIRCULATION OF

JOURNAL OF EXPERIMENTAL EDUCATION published QUARTERLY
(Insert exact title of publication) (State exact frequency of issue)

at Madison, Wisconsin for October 1, 1962
(Name of post office and State where publication has second-class entry)

1. The names and addresses of the publisher, editor, managing editor, and business managers are:

	Name	Address
Publisher	Dembar Educational Research Services, Inc.	Madison, Wisconsin
Editor	—	
Managing editor	Jean W. Williams	
Business manager	Arnold A. Caucutt	

2. The owner is: (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual member, must be given.)

Name	Address
Dembar Educational Research Services, Inc.	Madison, Wisconsin
Walter A. Frautschi	Madison, Wisconsin
Dorothy Jones Frautschi	Madison, Wisconsin
Mrs. A. S. Barr	Madison, Wisconsin

3. The known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities are: **(If there are none, so state.)**

Name	Address

4. Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting; also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner.

5. The average number of copies of each issue of this publication sold or distributed, through the mails or otherwise, to paid subscribers during the 12 months preceding the date shown above was: (This information is required by the act of June 11, 1960 to be included in all statements regardless of frequency of issue.) 1,149

Arnold A. Caucutt, Business Mgr.
(Signature of editor, publisher, business manager, or owner)

Sworn to and subscribed before me this 1st day of October, 1962

[SEAL]

Lorna Canfield

My Commission Expires July 26, 1964

(My commission expires ...., 19.....)

POD Form 3526
(July 1960)

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

PEER-SELECTION VS. EXPERT JUDGMENT AS A MEANS OF VALIDATING A TEACHER MORALE MEASURING INSTRUMENT

RALPH R. BENTLEY and AVERNO M. REMPEL
Purdue University

Purpose of the Study

Several attempts have been made to develop an instrument to measure teacher morale. Generally these instruments seem to have unknown validity and reliability.

The purpose of this study was to construct a teacher morale measuring instrument and to compare two approaches in establishing validity for the instrument. One method involved an analysis of item responses of "high" and "low" morale teachers identified on the basis of peer selections. The other approach was to use experts in identifying the responses to items which would be characteristic of those made by teachers with high morale.

Experimental Form of Morale Instrument

For purposes of this project, teacher morale was conceptually defined as "the professional interest and enthusiasm that a teacher displays toward the achievement of individual and group goals in a given school situation." Operationally, teacher morale was defined by the responses to the morale instrument. The morale instrument is referred to as a Teacher Opinionnaire.

The experimental form of the Teacher Opinionnaire consisted of 169 items based on other research instruments and the literature of the relevant social sciences—education, psychology, and sociology. Some 50 competent judges—teachers and psychologists—judged each item for clarity of meaning and degree of relevance. Items judged "lacking" or "low" in relevance, or on which there was not substantial agreement were eliminated. Of the 169 items 157 were retained to sample reliably eight categories: self-status, relationships with students, relationships with other teachers, factors relating to administration and policies, relationships with community, curriculum factors, working conditions and economic factors.

The items in the experimental form of the Teacher Opinionaire were randomly arranged. Sample items and the directions for responding are given below:

The purpose of this instrument is to obtain your opinions about your work as a teacher and about various school problems in your particular school situation.

Directions: Read each statement carefully. Then indicate whether you agree, probably agree, probably disagree, or disagree with each statement.

Use the following scale in recording your responses:

Circle "A" if you agree with the statement.

Circle "PA" if you are somewhat uncertain, but probably agree with the statement.

Circle "PD" if you are somewhat uncertain, but probably disagree with the statement.

Circle "D" if you disagree with the statement.

This opinionaire has no right or wrong responses, so do not hesitate to mark the statements frankly. Your answers will be treated as strictly confidential. DO NOT OMIT ANY ITEMS.

16. The faculty feels that the administration adequately represents and transmits the teachers' suggestions and interests to the board of education.................... A PA PD D

126. Vested interests of the faculty hinder the attainment of the educational objectives of our school................. A PA PD D

57. Our teaching staff is congenial to work with............ A PA PD D

52. The teachers in our school experience a sense of achievement and satisfaction in teaching their students .... A PA PD D

69. Lack of coordination at the administrative level unnecessarily complicates the work of the teachers........ A PA PD D

Peer Selections

One approach to validating the Teacher Opinionaire involved the identification of "high" and "low" morale teachers on the basis of peer selections. To accomplish this a Teacher Morale Rating Form was attached to the Teacher Opinionaire. In a particular school system, teachers were asked to identify by name on the Rating Form, depending on the size of the faculty, from three to ten teachers whom they considered to have the highest morale, and also to select an equal number whom they considered to have the lowest morale. Teachers were asked to use the conceptual definition of morale given previously, in making their judgments.

The other validating procedure involved the development of a scoring key based on expert opinions. This was accomplished by having a group of eight experts in measurement and school administration select independently the response for each item which they though would be made by high morale teachers. There was practically unanimous agreement in the judgments made.

A sample of 22 high schools was selected within a radius of one hundred miles from Purdue University. Faculties ranged in size from 10 to 58 teachers. Schools were selected so as to include both township and city administrative units. Administrators were contacted by members of the research team to determine their willingness to cooperate in the study. In some cases administrators deemed it necessary to gain the consent of the faculty.

Administration of the Instruments

Because of the nature of the responses and judgments which the teachers were asked to make, it was deemed essential to arrange for members of the research team to meet with the individual faculties of the cooperating schools to explain the nature and purposes of the study. After some preliminary tryouts, it was also found necessary to have the faculty complete the instruments during the time of the meetings. This procedure assured almost complete returns.

To encourage candid responses, teachers were instructed not to sign their names. However, for some of the analyses to be made it was necessary to identify the responses of particular teachers. So all forms were coded. This procedure was frankly discussed with each faculty and assurance was given that all data collected would be used and reported for research purposes only.

A uniform method of presentation was developed for the administration and completion of the Teacher Opinionaire and the Teacher Morale Rating Form.

Analysis of Data

The first step was to identify "high", "middle", and "low" teacher morale groups on the basis of the peer judgments. Teachers were assigned to either the "high" or "low" groups if the ratio of the difference between the number of high and low selections received to the number of possible selections was approximately .25 or higher. Those not included in either the "high" or "low" categories were considered to constitute the "middle" group.

A stratified random sample of 100 teachers was selected to serve as a hold-out sample to be used in computing a validity coefficient. The sample selected was proportional with respect to size of school faculty and to the number of teachers in the "high", "middle", and "low" groups. The remaining 378 teachers were used for item analysis purposes.

Discrimination and difficulty indices were computed for each item on the basis of the percentages of the "high" and "low" morale (peer selections) groups checking the various item responses, using a table[1] of item-difficulty and item-discrimination indices for given proportions of success in the highest 27 percent and the lowest 27 percent of a normal bivariate population. Items were selected for cross-validation purposes on the basis of their discriminating power and their level of difficulty, and a scoring key was developed using only these items.

The scoring key was used to determine the morale scores for the hold-out sample. The significance of the differences between means of the "high", "middle", and the "low" groups was tested by analysis of variance procedures. The correlation ratio (usually referred to as the eta coefficient) was then computed.[2] This ratio was considered to be a measure of validity based on peer selections.

A scoring key was developed based on the opinions of eight experts as to how "high" morale teachers would respond to each item. Then, for the teachers not in the hold-out sample, total morale and part scores for each of the eight categories of the opinionaire were obtained, using the scoring key based on expert judgments.

Individuals (hold-out group not included) with total scores in the upper 27 percent of the distribution were assigned to the "high" morale group, and those with scores in the lower 27 percent of the distribution were placed in the "low" morale group. Discrimination and difficulty indices were computed using the item analysis table previously cited. Items were retained on the basis of their discriminating power and level of difficulty.

Using the same hold-out group described earlier, Kuder-Richardson estimates (K-R 21) of reliability[3] were determined for the total score and

TABLE I

NUMBER AND PERCENTAGE OF TEACHERS IDENTIFIED BY PEERS WHO WERE INCLUDED IN THE "HIGH" AND "LOW" MORALE GROUPS BY SCHOOLS

School	Total Number Teachers in School	High Morale Group		Low Morale Group	
		Number	Percentage	Number	Percentage
1	58	10	17	8	14
2	49	7	14	9	20
3	48	13	27	8	17
4	39	12	31	8	21
5	34	9	26	7	21
6	33	9	27	7	21
7	31	11	35	6	19
8	26	7	27	8	31
9	26	7	27	8	31
10	24	7	29	8	39
11	22	6	27	7	32
12	22	5	23	6	27
13	20	7	35	6	30
14	19	7	41	6	35
15	19	6	32	6	32
16	18	4	22	5	28
17	15	5	33	5	33
18	15	7	47	5	33
19	14	6	43	5	36
20	14	6	43	4	29
21	14	6	43	5	36
22	10	3	30	3	30
Total	570	160	28	140	25

TABLE II

GROUP FREQUENCIES OF ITEM DISCRIMINATION AND ITEM DIFFICULTY INDICES RESULTING FROM COMPARISON OF RESPONSES OF PEER-SELECTED "HIGH" AND "LOW" MORALE GROUPS

Discrimination Indices	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
90-99										
80-89										
70-79										
60-69										
50-59										
40-49										
30-39					1		2			
20-29		1		1	1	3	3	4	3	2
10-19	2	6	8	7	13	10	17	9	5	2
0-9	4	5	2	7	8	10	6	7	7	1
	Difficulty Indices									

TABLE III

GROUPED FREQUENCIES OF ITEM DISCRIMINATION AND ITEM DIFFICULTY INDICES BASED ON COMPARISON OF "HIGH" AND "LOW" MORALE GROUPS IDENTIFIED ON THE BASIS OF TOTAL SCORES USING THE EXPERT KEY

Discrimination Indices	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
90-99										
80-89					1					
70-79			6	1	2	4	3	2		
60-69		2	6	16	9	11	6	2	1	
50-59		4	6	11	8	11	11	2	2	
40-49	2	1	2	2	5	1	4	2		
30-39		1	1	2	1	3	2	1		
20-29										
10-19										
0-9										
	Difficulty Indices									

category scores; also, intercorrelations were calculated between the category scores and then corrected for attenuation.

Mean teacher morale scores and standard deviations were computed for each of the 22 schools. Analysis of variance procedures were used to determine whether the school morale means differed significantly.

Results

Table I indicates the total numbers of teachers in each of the 22 schools, and the number and percentage of teachers identified by their peers and included in the "high" and "low" morale groups in each school.

The distribution of item discrimination and item difficulty indices using peer selections for determining extreme groups is given in Table II. Discrimination indices ranged from .00 to .38; difficulty indices from .06 to .95.

Ninety-three (93) items having discriminating indices of .10 or more and difficulty indices ranging from .10 to .90 were retained for cross-validation purposes.

Mean opinionaire scores for the "high", "middle", and "low" morale groups (peer selections) in the hold-out sample were 54, 53, and 46, respectively. Using analysis of variance procedures, the obtained F value of 1.90 with d.f. = 2 and 99 is not significant at the 5 percent level of significance. The eta coefficient was .19 and also not significant. Even though differences among the means are not significant, it should be noted that the trend of the scores is in the expected direction.

Table III presents the distribution of item discriminations and difficulty indices computed by comparing responses of the "high" and "low" morale groups identified on the basis of total scores using the expert key. The range of discrimination indices was from .30 to .80; the range of difficulty indices was from .06 to .85.

Most of the items with discrimination indices below .40 were eliminated; however, in a few cases they were retained if the item fell in a content category having a limited number of items. Generally, items with difficulty indices below .20 and above .80 were also eliminated. Altogether, 146 of the 157 items were retained.

The Kuder-Richardson reliability coefficient for total scores using the hold-out sample was .96. Reliability coefficients for category scores ranged from .79 for economic factors to .98 for factors relating to administration and policies. (See Table IV.)

In Table V, intercorrelations between category scores are listed. Intercorrelations are also given when corrected for attenuation. The highest corrected correlation (.74) was found between self-status and relationships with students. Lowest corrected correlations (all .47) were obtained between self-status and relationships with community, between relationships with students and relationships with community, and between economic factors and relationships with students. All intercorrelations were well above .26, the value needed to be significant at the one percent level.

As can be seen from Table VI, mean teacher morale scores ranged from 33.75 to 70.47; standard deviation values varied from 18.71 to 36.42. The F ratio resulting from the analysis of variance with 21 and 487 degrees of freedom was 1.98, statistically significant beyond the one percent level. (A Bartlett Test of variance homogeneity preceded the analysis of variance and resulted in a non-significant chi-square.)

As a final step in the analysis, mean teacher questionnaire scores were computed for the peer-selected "high", "middle", and "low" teacher morale groups and tests of significance made of the differences between the means. This was a way of checking how well the 146-item instrument, using a scoring key developed on the basis of expert judgment, discriminated between "high", "middle", and "low" morale groups identified by means of another criterion, i.e., peer selection.

Table VII presents the means and standard deviations of total opinionaire morale scores of peer-selected "high", "middle", and "low" morale groups for the hold-out sample, the item analysis sample, and also for the total sample of teachers. It will be observed that the corresponding means of the "high", "middle", and "low" groups are practically identical for the three samples of teachers. The F-ratio (d.f. = 2 and 477) resulting from the analysis of variance of the means of the total sample was 3.24, significant at the five percent level.

Summary and Conclusions

A teacher morale measuring instrument consisting of 157 items in eight major categories was administered to 570 teachers in 22 Indiana high schools. Two methods of establishing validity for the instrument were studied. One approach involved using the criterion of peer identification of "high" and "low" morale teachers. The other method used the criterion of expert judgment as to how items would be scored by "high" morale teachers.

The major findings are:

1. A large majority of the items in the teacher morale measuring instrument showed low discriminating power between "high" and "low" morale teachers as identified by their peers. The validity coefficient obtained for a hold-out group of 99 teachers, using 93 of the most highly discriminating items, was not significant.

2. Using a scoring key based on expert judgment, practically all of the items adequately discriminated between teachers with "high" and "low" total scores. Discrimination indices ranged from

TABLE IV

INTERNAL CONSISTENCY RELIABILITY COEFFICIENTS FOR TOTAL SCORE AND FOR CATEGORY SCORES

	Category	Number of Items	Reliability Coefficients
1.	Self-status	24	.93
2.	Relationships with Students	10	.86
3.	Relationships with Other Teachers	19	.93
4.	Factors Relating to Administration and Policies	51	.98
5.	Relationships with Community	11	.82
6.	Curriculum Factors	11	.90
7.	Working Conditions	9	.84
8.	Economic Factors	11	.79
	Total	146	.96

TABLE V

INTERCORRELATIONS BETWEEN CATEGORY SCORES

	1	2	3	4	5	6	7
2	.66 (.74)*						
3	.47 (.51)	.58 (.65)					
4	.57 (.60)	.49 (.53)	.63 (.66)				
5	.41 (.47)	.40 (.47)	.59 (.68)	.64 (.71)			
6	.59 (.64)	.55 (.62)	.55 (.60)	.59 (.63)	.51 (.59)		
7	.50 (.57)	.43 (.50)	.42 (.48)	.64 (.70)	.50 (.60)	.52 (.59)	
8	.46 (.54)	.39 (.47)	.45 (.53)	.60 (.68)	.44 (.54)	.59 (.70)	.55 (.67)

1% level of significance .26
(N = 99)

1. Self-status
2. Relationships with students
3. Relationships with other teachers
4. Factors relating to administration and policies
5. Relationships with community
6. Curriculum factors
7. Working conditions
8. Economic factors

*Correlation values shown in parentheses are corrected for attenuation.

TABLE VI

MEAN TEACHER MORALE SCORES AND STANDARD DEVIATIONS BY SCHOOLS

School Code	No. Teachers Responding	Mean	Standard Deviation
19	19	70.47	35.64
08	21	68.24	36.21
04	46	66.28	34.23
05	18	66.33	36.42
17	12	65.67	34.00
03	13	64.69	35.26
10	25	62.92	36.00
20	20	62.85	36.00
02	26	59.85	24.48
16	42	58.95	27.35
22	17	58.71	33.46
12	14	57.64	29.60
11	23	57.74	25.77
14	33	57.55	33.00
01	46	55.04	26.25
13	10	54.60	26.51
07	29	52.34	33.88
15	18	49.67	23.37
21	33	48.78	27.19
09	13	38.69	18.71
18	23	35.74	23.37
06	8	33.75	25.89

TABLE VII

MEANS AND STANDARD DEVIATIONS OF TOTAL MORALE SCORES OF PEER-SELECTED "HIGH", "MIDDLE", AND "LOW" TEACHER MORALE GROUPS FOR HOLD-OUT ITEM ANALYSIS AND TOTAL SAMPLES OF TEACHERS

	Hold-out Sample			Item Analysis Sample			Total Sample		
	No.*	Mean	S. D.	No.	Mean	S. D.	No.	Mean	S. D.
High	25	68	21	92	67	30	115	67	28
Middle	50	65	24	198	65	30	246	65	29
Low	24	59	34	96	58	29	117	58	30

*One teacher was dropped from the hold-out sample because of faulty coding.

.30 to .80. The instrument was reduced to 146 items.

a. The 146-item instrument yielded an internal consistency reliability coefficient of .96. Internal consistency reliability coefficients for the various categories taken separately ranged from .79 to .98.

b. All intercorrelations between category scores were significant at the one percent level and varied from .47 to .74 when corrected for attenuation.

c. On the basis of total scores highly significant differences were found in the average level of teacher morale among the 22 schools. Mean scores by schools ranged from 33.75 to 70.47.

3. To study the relationship between the two validating criteria, means of the total scores (using the expert judgment scoring key) were compared for the peer-selected "high", "middle", and "low", morale groups, using the total sample of 478 teachers. The F-ratio resulting from the analysis of variance was significant at the five percent level. The pattern of means obtained with the 146-item opinionaire for the "high", "middle", and "low" groups (67, 65, and 58), using the total sample of teachers, was identical with the pattern of means obtained (54, 53, and 46) with the 93-item opinionaire using the hold-out sample. The fact that the obtained F value was significant in the first instance and not significant in the second can be attributed to the larger N (478 vs., 99).

In comparing the two approaches in the validation of the instrument, the findings do not provide an unequivocal answer. The larger discrimination indices obtained for the expert judgment—internal consistency procedures is as expected and agrees with most research results dealing with the construction of tests and attitude scales. However, the significant difference that was obtained in the means for the total sample between peer-selected "high", "middle", and "low" morale groups using the expert scoring key is encouraging. It indicates that there is considerable relationship between the two criteria.

One limitation of the peer selection procedure should be pointed out. It was found in this study that there were marked differences in the level of morale between schools. Since peers judged relative morale only in terms of their particular school, there is probably some overlapping between "high", and "low" morale groups in different schools, i.e., "low" morale teachers in one school may be at the level of "high" morale teachers in another school.

The 146-item opinionaire (expert judgment) has one distinct advantage over the 93-item (peer-selection) opinionaire. The larger number of items in the former makes it possible to obtain category scores for each of eight selected categories. Reducing the instrument to 93 items would necessitate collapsing some of the categories.

The various category scores do not correlate so highly with each other, even when corrected for attentuation, that it can be presumed that they measure entirely the same factors in the morale complex. Yet they are fairly reliable in an internal consistency sense. These features in combination with the fact that differences between mean scores for schools were highly significant, would indicate that a valuable diagnostic instrument has been developed.

FOOTNOTES

1. Chung-Teh, Fan. Item Analysis Table (Princeton: Educational Testing Service, 1952).
2. J. P. Guilford. Fundamental Statistics in Psychology and Education (1950), pp. 314-23.
3. J. P. Guilford. Op. cit., pp. 494-98.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

AN ANALYSIS OF INTERACTIVE BEHAVIOR IN A PRACTICUM IN PERSONNEL RELATIONSHIPS*

WILLIAM C. STRASSER, JR.
State University of New York at Buffalo

Problem Investigated

Purpose and Rationale. The purpose of this research is to examine the validity of the general hypothesis that: patterns of operational behavior of persons participating in a Practicum in Personnel Relationships can be identified, the patterns being relevant to attitude, flow of discussion, interaction during discussion, and seating contiguity. Analysis of the behavior of participants while they are engaged in sessions of a practicum would provide appropriate data for an evaluation of the practicum. If it be possible to discern patterns of operational behavior in a human relation training program like the Practicum in Personnel Relationships, these patterns could be related to appropriate criteria and a relatively controlled, objective assessment of the program could be made in terms of operational behavior rather than, or as well as, in terms of written reports or self-analyses.

Though several methods of analysis reported herein have been used by various investigators, three relatively new methods of analysis were developed by the investigator for this research on the Practicum: thread-flow diagrams of participation in group discussion, an analysis of the equitableness of the distribution of participation in group discussion, and a seating contiguity index. If the methods of analysis used for this study be effective in identifying patterns of operational behavior, the methods could be used in other research studies on human relations training programs.

The research design is a developmental, longitudinal survey of the four behavioral variables enumerated in the general hypothesis above.

Review of the Literature. Two supporting concepts are found in the literature on human relations training. The first concept is that human relations training is an important aspect of programs for the preparation and development of teachers and educational administrators. The second concept is that there is a continuing need for research regarding the analysis and evaluation of human relations training. Bonner (3:327), Moore (6:28-9), and numerous abstracts in Dissertation Abstracts, in the subdivision on "Education," attest to the importance and prevalence of human relations training for educators. Hughes (4:14) and Miles (5:252) point out the need for research and reports on human relations training programs. That the American Educational Research Association publishes a volume on "Human Relations in Education", (1) is testimony to the concern of educators for research in this field. Insofar as this investigator knows, this report on the Practicum is the first extensive research dealing with observations of these selected networks of interactive behavior as a means of analyzing a program in human relations. Even though the findings have limited generalizability, it is hoped that the research will be helpful in the designing of methods and the testing of hypotheses by other investigators concerned with human relations training.

The Practicum. This report is extracted from an extensive analysis and evaluation of the Practicum in Personnel Relationships, which is a graduate program in human relations for school personnel and is conducted annually at the University of Maryland (9). The Practicum is designed to provide for school personnel who have completed at least several years of successful teaching experience, an inductive, laboratory-like experience in personnel relations (7). Dr. Clarence A. Newell, professor of educational administration of the University of Maryland, developed and instructs the Practicum, which has been offered each academic year since 1948. School personnel who express a desire to participate in the Practicum are selected each fall to be members of a group which studies a

*This article is based upon a Ph. D. dissertation submitted to the University of Maryland and was delivered at the annual meeting of the American Educational Research Association, 1961.

philosophy of democratic human relations and practices selected techniques in group processes and person-to-person relations. The techniques studied are useful in school administrative and teaching situations. The main teaching method in the Practicum is group discussion. The group formulates its own goals, plans its own activities, executes its plans, and evaluates its effectiveness in achieving its goals. The instructor participates as a member of the group. The group meets for two and on-half hours a week for an entire academic year, September through May. Membership in a group is limited to fifteen through twenty persons.

Subjects. The subjects studied in this research were members of two groups of the Practicum for 1959-60. Group One met in College Park, Maryland, and comprised 16 persons plus the instructor. Group Two met in Baltimore, Maryland, and comprised 18 members plus the instructor. Group One was composed of 14 males and two females. Group Two was composed of 13 males and five females. Chi square computations revealed no statistically significant differences, five percent level, between the two groups with respect to the following selected personal characteristics: sex, age, years of teaching experience, years of administrative experience in education, and type of position currently held.

Definition of Variables Investigated. Four interactive variables were investigated. The first variable, attitude, is defined as the numerical responses of the subjects to a structured personnel-relationships attitude questionnaire designed by the investigator and articulated with selected criteria of democratic education leadership (see paragraph below). The second variable, flow of discussion, is defined as the sequence of persons entering into verbal discussion during selected meetings of the two groups. The third variable, interaction during discussion, is defined as the overt responses of the subjects during selected group discussions, as classified in the Bales interaction process analysis (2:177-95). The fourth variable, seating contiguity, is defined as the extent to which a subject sits next to different subjects during selected training sessions, and is represented as the ratio of a) the number of different subjects next to whom a given subject sat during each of a given number of consecutive meetings, to b) the potential number of different subjects next to whom he could have sat during those consecutive meetings. (Subjects sat in a circular-type seating pattern at each meeting.)

Criteria of Democratic Education Leadership. In order to provide a frame of reference for the attitude questionnaire and for the analysis of data, the literature on democratic education leadership was reviewed and the following operational criteria were developed:

In his general operational approach, a democratic education leader provides for:

1. Cooperative development of goals, realizing that administration is a service function, not an end in itself or an arbitrary authority.
2. An organization of the staff that will make possible the participation of the staff and the cooperation of the staff and the community in executing the education function.
3. Cooperative evaluation of actions and achievements, with reference to specific goals cooperatively evolved.
4. Research based upon scientific principles, to the end that methods for participation, execution, and evaluation are improved and adapted to changing situations.
5. Utilization of cooperative means in the classroom, and among staff, citizens and pupils.
6. Coordination of school goals and activities with community goals and activities, inasmuch as the school is but an institution of the community and should be articulated with the community.
7. Development of the concept that an individual is dependent upon groups and the community for the satisfaction of his personal needs.

In discussion and interaction situations, the democratic education leader provides for:

1. Participation of all members of the group in the solution of group problems and achievement of group goals.
2. Achievement of group decision, preferably through the development of consensus rather than through majority rule or through minority manipulation of the majority.
3. Discussion interaction with as many members as is possible, all members optimally.
4. Discussion interaction in differenct roles, as need arises.
5. Meetings in which a person may assume a seat beside as many different members as possible, from meeting to meeting, to facilitate interaction.

Procedures

General Procedures for Collection of Data. The data on operational behavior were gathered independently for each group, the investigator observing each group fourteen times during the period September 1959 through April 1960. During observations the investigator usually sat outside of the circular seating arrangement which typically is used during meetings. He could see all subjects in the group and could hear all audible group discussion. He did not enter into group discussions except when administering a form, or when asking for the permission of the group to carry out a particular action. Selected discussions were tape recorded to provide a basis for checking the reliability of the observations made with a modified Bales interaction system. Data were collected and analyzed with reference to each of the four variables,

as explained below.

Procedures Related to Attitude. Attitude was measured twice during the observations, once in September at the first meeting of each group and once again for each group during a meeting in May 1960. A twenty-six item questionnaire on personnel relationships, especially related to public school situations, was developed by the investigator and administered to individual subjects as a group (9:191-5). Each subject responded to each item, indicating by a written numerical symbol (± 1, 2, 3) the extent of his agreement (+) or disagreement (-) with the item. Each response was translated into a positive, scaled number, whose value depended upon its agreement or disagreement with the criteria of democratic education leadership, above. A total score was determined for each subject, for each of the two administrations. Correlated scores for individual subjects – September score compared with May score – were analyzed for the significance of differences in each group. The Wilcoxon Test for Paired Replicates, one-tailed, was used for this analysis (8:489-91). The five percent level of significance was established for satisfactory statistical significance.

Procedures Related to Flow of Discussion. Flow of discussion was observed during seven meetings of each group from September through April. The investigator collected data on the flow of discussion by noting on a special flow chart the sequence of subjects entering into the group discussion during a period of one to two hours. As each subject spoke, the investigator noted the alphabetical symbol (e.g., "A" or "B" etc.) representing that subject, the symbols being recorded in sequence from left to right, thus: A B J E K S, etc. After each such observation of a meeting, the recorded flow sequence was analyzed in terms of two sub-variables: quantity of participation per subject per meeting and a thread-flow reproduction of each discussion session. The quantity of discussion per subject was determined by counting the number of symbols for the meeting and comparing this number with the total number of symbols for the meeting. The comparison was expressed as a decimal proportion per subject. By means of the index "relative uncertainty," the equitableness of the distribution of participation among subjects present was calculated (i.e., the extent to which all subjects participated equally in the group discussion). The "relative uncertainty" formula is:

$$H_{rel} = \frac{H}{H_{max}}, \text{ where } H = -\sum_{1}^{n} p\,(i) \log_2 p\,(i), \text{ and}$$

$$H_{max} = -\sum_{1}^{n} \frac{1}{n} \log_2 \frac{1}{n}.$$

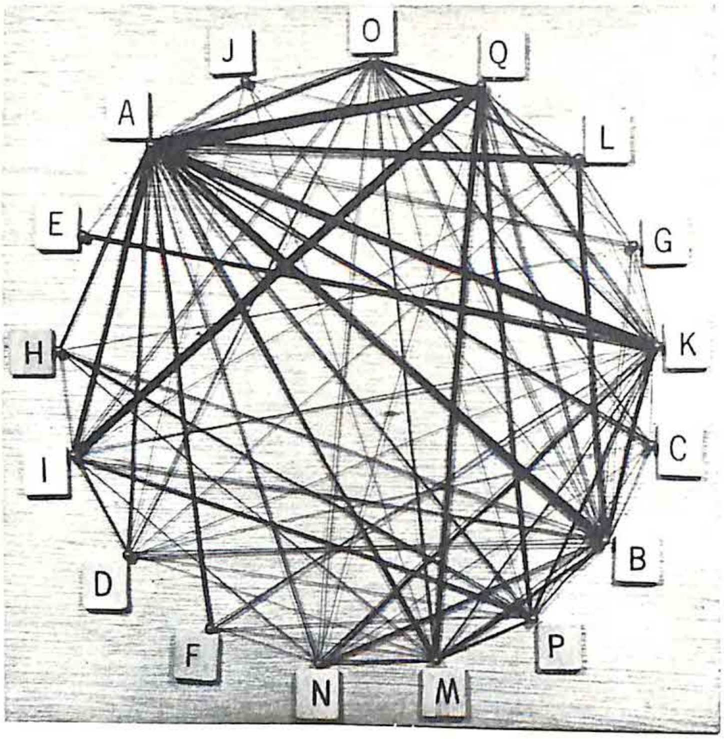
(See bibliographic entry 8:78-99, for an explanation of this index.) In the context of the flow of discussion, "p (i)" is the proportion of participation for each subject present, and "n" is the number of subjects present at a given meeting. The numerical value of "relative uncertainty" ranges between positive one and zero. The greater the numerical value of this index is, the more equitably distributed is the amount of participation per subject present.

Additionally, the data on the flow of discussion were analyzed by means of thread-flow diagrams, which reproduced the flow network for each meeting observed. The investigator devised a technique whereby the flow of discussion could be diagrammed on a wooden board on which were placed map pins to mark the seating position of each subject and across which was woven from pin to pin a black sewing thread in the same sequence as indicated on the flow chart recorded by the investigator. The next page contains photographs of two of these thread-flow diagrams. The diagrams were examined to investigate the changes visually apparent in the flow networks of the two groups during the academic year.

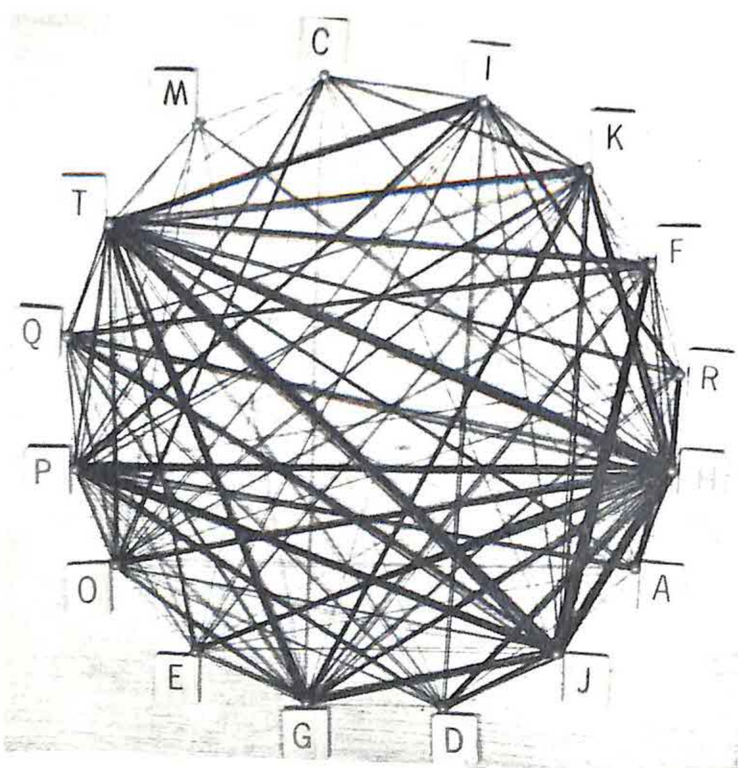
Procedures Related to Interaction During Discussion. Interaction during discussion was observed at seven meetings of each group during the academic year. A modified Bales interaction analysis system was used, in which each subject's verbal and nonverbal behavior during selected group discussions was classified in terms of twelve categories of discussion interaction. These twelve categories are: shows solidarity, shows tension release, agrees, gives suggestions or direction, gives opinion or evaluation, gives orientation or information, asks for orientation or information, asks for opinion or evaluation, asks for suggestion or direction, disagrees, shows tension, and shows antagonism. As each subject spoke, the investigator analyzed the subject's comments in accordance with these categories and recorded the subject's alphabetical symbol on a special chart. Nonverbal behavior was also analyzed - such as nodding of heads, taking notes, smiling, and withdrawing from discussion to engage in a private conversation. Thus, a "profile" of the quality of interaction was developed for each group for each of the seven meetings and for each subject for these meetings. The profiles of each group were examined to analyze the patterns of the quality of interaction in each group during the academic year. The analysis of each subject's profile, or "role," was carried out by means of the index "relative uncertainty," explained above. However, for the analysis of a subject's role,

$$H = -\sum_{1}^{12} p\,(i) \log_a p\,(i), \text{ and } H_{max} = -\sum_{1}^{12} \log_a \left(\frac{1}{12}\right),$$

where "p (i)" is the proportion of interaction units of a given subject in each of the twelve Bales categories, taken separately, and "$\frac{1}{12}$" represents the ideal proportion of interaction units per Bales category if there be completely equitable distribution among all categories. The higher the



AN ANALYSIS OF INTERACTIVE BEHAVIOR IN A PRACTICUM
IN PERSONNEL RELATIONSHIPS: APPENDIX A
GROUP II, OCTOBER 14, 1959



AN ANALYSIS OF INTERACTIVE BEHAVIOR IN A PRACTICUM
IN PERSONNEL RELATIONSHIPS: APPENDIX A
GROUP II, APRIL 20, 1960

numerical value of the "relative uncertainty" index is, the greater is the equitableness of roles manifested by a given subject during a given meeting.

Procedures Related to Seating Contiguity. Seating contiguity was observed at the first six consecutive meetings of each group and at six consecutive meetings in the spring of 1960. The investigator gathered data on seating contiguity by recording the seating arrangement for each of these meetings, using the alphabetical symbol representing each subject. The investigator devised a "seating contiguity index" (SCI) to obtain a statistical measure of the extent to which a subject sat next to different subjects during two different periods of time. The formula devised is:

$$SCI = \frac{\sum^{N} DM}{\frac{1}{2N}}$$

where "DM" is "the number of different subjects next to whom a given subject sat during a given meeting," and "N" is the "number of consecutive meetings considered in developing the index." The SCI for this research varied between positive one and zero. The higher the numerical value of SCI is, the greater is the proportion of different subjects next to whom a subject sat during the meetings. Two SCI's were derived for each subject, one for the fall interval and one for the spring interval. These correlated indices were compared to determine whether there was a numerical increase in the variation of seating contiguity in each group, fall compared with spring. The Sign Test, one-tailed, was used to determine the statistical significance of the correlated differences of SCI's in each group (8:441-3). The five percent level of significance was established as satisfactory.

Findings

Findings Related to Attitude. The potential raw score range on the personnel-relationships attitude questionnaire was 26 through 182 points. The raw score range attained by subjects in Group One was 136 through 176 points for the first administration, with a median score of 156; on the second administration, the range was 134 through 173, and the median 158. The raw score range for subjects in Group Two was 145 through 172 for the first administration, with a median of 162; on the second administration, the range was 135 through 182, with a median of 164. Thus, the pattern of scores indicates a relatively high agreement with the criteria of democratic education leadership.

An analysis of the increase and decrease of raw scores by the Wilcoxon Paired Replicates Test indicates that the change in correlated scores was slightly in the direction of increased agreement with the criteria of democratic education leadership. However, the change in scores was not statistically significant at the five percent level.

Findings Related to the Flow of Discussion. The pattern of quantity, or proportion, of discussion per subject per meeting observed for flow was generally the same for both groups Table I illustrates the proportion of flow units per subject in Group One and the corresponding "relative uncertainty" index per meeting. In Group One, the least proportion per subject ranged between .000 and .009 for the seven meetings, with a median least-proportion of .002; the median proportion ranged between .044 and .064, with a median median-proportion of .049; the greatest proportion ranged between .148 and .209, with a median greatest-proportion of .175. In group Two, the least proportion per subject ranged between .000 and .012, with a median least-proportion of .006; the median proportion ranged between .024 and .056, with a median median-proportion of .039; the greatest proportion ranged between .156 and .287, with a median greatest-proportion of .207.

The "relative uncertainty" index indicates a pattern of slight improvement in each group in the equitableness of the distribution of proportions among subjects present during the seven selected meetings observed for flow. For Group One, the mean value of the index for the first semester of the academic year is .874, and for the second semester, .904. For Group Two, the mean value for the first semester is .865, and for the second semester, .873.

The thread-flow diagrams revealed five patterns in the flow of discussion. First, with the exception of one subject at each of two meetings, the flow of discussion involved every subject present at the meetings represented by the flow diagrams. Second, as time passed, participation leadership in each group (i.e., the heaviness of the black-thread lines emanating from a pin) tended to be dispersed among a relatively increasing number of subjects. Third, as time passed, discussion tended to be interchanged among an increasing number of subjects (i.e., there tended to be an increasing number of heavy black lines between pins). Fourth, as time passed, the subjects who served as chairmen (the chairmanship rotating among all the subjects from meeting to meeting during the year) tended to be less and less dominant as focal subjects for the flow of discussion. Fifth, as time passed, the instructor (letter Q in the diagrams) tended to disperse his discussion among a relatively increasing number of subjects (i.e., there were more thread lines emanating from the pin representing him to other pins) and tended to be less and less dominant as a member of the discussion groups (i.e., the heaviness of the thread lines emanating from his pin became increasingly like the heaviness of other thread lines).

Findings Related to Interaction During Discussion. With reference to the twelve-category Bales classification of interaction, three general patterns of interaction were apparent in the two groups. First, there was a sustained variety of interaction

TABLE I

PROPORTION OF FLOW UNITS PER SUBJECT, GROUP ONE
By Date of Meeting

Subject	Date of Meeting						
	10-13	10-20	12-1	2-23	3-8	3-15	4-5
A	.209	.053	.020	.085	.105	.097	a*
B	.091	.193	.057	nr*	nr	nr	nr
C	.018	.057	.002	.040	.000	.048	.019
D	.022	.032	.020	067	.015	.011	.014
E	.010	.008	.030	.029	.049	.011	.028
F	.014	.015	.175	.036	.032	.148	.001
G	.012	.039	.060	nr	nr	nr	nr
H	.020	.061	.047	.040	.027	.029	a
I	.128	.031	.071	.074	.025	.012	a
J	.008	.011	.004	.009	.006	.044	.064
K	.097	.145	.165	.090	.143	.054	.189
L	.038	a	a	.047	.032	a	.109
M	.055	.084	.127	.061	.160	.148	.097
N	.030	.049	.068	.112	.065	.116	.100
O	.049	.023	a	.020	.086	.046	.125
P	.073	.036	.031	.159	.061	.001	.047
Q	.126	.163	.121	.058	.112	.092	.064
R	nr	nr	nr	.031	.072	.098	.115
S	nr	nr	nr	.045	.011	.045	.028
Least Proportion	.008	.008	.002	.009	.000	.001	.001
Median Proportion	.049	.044	.057	.047	.049	.047	.064
Greatest Proportion	.209	.193	.175	.159	.160	.148	.189
Relative Uncertainty Index	.865	.885	.872	.939	.882	.897	.897

*Key: The symbol nr denotes the subject was not registered in the Practicum. The symbol a denotes the subject was absent from the meeting.

among the twelve Bales categories throughout the observations during the academic year. (Interaction took place in all twelve categories during eleven of the fourteen meetings observed with the Bales: at the other three meetings interaction took place in eleven of the twelve categories.) Second, the proportion of subjects interacting in each category was relatively high throughout the period of observations. (More than half of the subjects interacted in more than half of the categories during the fourteen meetings observed.) Third, the subjects tended to interact in a relatively high number of Bales categories throughout the observations. (The median number of categories in which a subject interacted was 7 through 9 for twelve of the fourteen meetings observed: for the other two meetings, the median number was 4 through 6.)

The "relative uncertainty" index, as applied to the analysis of the equitableness of the distribution of interaction by a subject among the twelve Bales categories, indicates that there was a pattern of sustained and fairly high equitableness per subject during the fourteen observations. The least "relative uncertainty" index per meeting ranged between .123 and .576 during these fourteen meetings. The median index ranged between .663 and .748. The greatest index ranged between .777 and .851. The potential range is .000 to 1.000.

According to the method recommended by Bales to check the reliability of observations, the numerical reliability of the observations is generally satisfactory within a limit of a two-sigma deviation (2:110-2). The reliability of interaction ratings was checked for selected subjects and for each group for selected meetings. A tape recording was used as the basis for the determination of reliability. Though several categories deviated from the two-sigma limit, the deviation may be attributed generally to the nonreproduction of nonverbal interaction on the tape recordings.

Findings Related to the Correlation of Quality and Quantity of Interaction. A high, positive rank-order correlation was found between the proportion, or quantity, of interaction per subject per meeting and his "relative uncertainty" index, or quality of interaction, for that meeting. This suggests that the following quantity-quality pattern exists: the more interaction in which a subject engages, the greater tends to be the equitableness of his distribution of interaction among the twelve Bales categories. For Group One, this rank-order correlation ranged from positive .459 to positive .841, with a median of positive .661. For Group Two, the correlation ranged from positive .421 to positive .894, with a median of positive .686.

Findings Related to Seating Contiguity. The seating contiguity indices per subject – the fall indices compared with spring indices on a correlated basis – indicate the following seating pattern: subjects tended to sit next to more different subjects in the spring semester than they did in the fall semester. Table II illustrates the seating contiguity indices per subject in Group Two. In Group One, 10 subjects increased their indices and 4 subjects decreased their indices. In Group Two, 12 subjects increased their indices and 3 subjects decreased their indices. The increase of indices in Group Two was statistically significant at the five percent level, according to the Sign Test, one-tailed. The increase of indices in Group One was not statistically significant at this level.

Conclusions and Recommendations

Conclusions. With reference to the general hypothesis stated at the beginning of this paper, it is possible to identify patterns of operational behavior of persons participating in a Practicum in Personnel Relationships, the patterns being relevant especially to flow of discussion, interaction during discussion, and seating contiguity, and being somewhat less clearly relevant to attitude. Though almost all of the patterns discussed in the findings above are not statistically significant, they are discernible and relatable to a set of criteria such as the criteria for democratic education leadership. Hence, one could conclude that it is possible to evaluate a human relations training program such as the Practicum in terms of the manifest, operational behavior of participants as well as (or instead of) in terms of self reports or evaluations, or in terms of the results of written tests, or in terms of evaluations of research papers written by participants in the program.

Additionally, the data gathered in one human relations program by methods such as those used for this research could be compared with data gathered in another human relations program or in the same program over a period of years. In this manner, relative training outcomes could be assessed from program to program and within the same program from time to time. Inasmuch as the methods used in this research on the Practicum were effective, they could be applied in other research on human relations.

Recommendation. Further research should be conducted to test the validity of the findings on the patterns of operational behavior discussed above, with reference to programs like the Practicum. In this connection, the hypotheses reported in the evaluation of the Practicum could be investigated (9:180-3). The research could be either experimental or survey.

Furthermore, the behavioral patterns reported herein should be investigated with a view toward identifying possible causal relationships. The identification of such relationships may not be conclusive, but may stimulate additional study and research.

Finally, there should be additional research directed toward the identification of other operational behavior patterns than those reported herein.

TABLE II

CONTIGUITY INDICES PER SUBJECT IN GROUP TWO, WITH ANALYSIS OF STATISTICAL SIGNIFICANCE OF INCREASE OF INDICES

By Fall and Spring Indices

Subject	Fall Index	Spring Index	Increase or Decrease
A	.583	.750	Increase
B	.600	nr*	na*
C	.583	.500	Decrease
D	.583	.750	Increase
E	.667	.600	Decrease
F	.500	.600	Increase
G	.583	.600	Increase
H	.333	.500	Increase
I	.667	.750	Increase
J	.417	.625	Increase
K	.750	.600	Decrease
L	.500	.875	Increase
M	.583	.583	Same
N	.500	.800	Increase
O	.583	.750	Increase
P	.417	.667	Increase
Q	.583	.833	Increase
R	nr	.600	na
S	nr	.667	na
T	nr	.600	na
Median Index	.583	.625	
	Number who increased contiguity index		12
	Number who decreased contiguity index		3
	Null hypothesis rejected		

*The symbol nr denotes the subject was not registered. The symbol na denotes the comparison is not applicable.

REFERENCES

1. American Educational Research Association, "Human Relations in Education," Review of Educational Research, Vol. XXIX No. 4 (Oct 1959), Washington, D. C.: National Education Association of the United States.
2. Bales, R. F., Interaction Process Analysis, Cambridge, Massachusetts: Addison-Wesley Press, Inc., 1950.
3. Bonner, H., Group Dynamics, Principles and Applications, New York: Ronald Press Co. 1959.
4. Hughes, J. M., Human Relations in Educational Organization, New York: Harper Brothers, 1957.
5. Miles, M. B., Learning to Work in Groups: A Program Guide for Education Leaders, New York: Teachers College, Columbia University, 1959.
6. Moore, H. A., Jr., Studies in School Administration, Washington, D. C.: American Association of School Adminisistrators, 1957.
7. Newell, C. A., "Administrators Learn Democratic Behavior," Educational Leadership, Vol. XI (1953), pp. 114-7.
8. Senders, V. L., Measurement and Statistics, New York: Oxford University Press, 1958.
9. Strasser, W. C., Jr., An Analysis and Evaluation of the Practicum in Personnel Relationships, unpublished doctoral dissertation (copyright), University of Maryland, 1961.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

A COMPARATIVE INVESTIGATION OF TWO METHODS OF TEACHING PHONICS IN A MODERN READING PROGRAM: A PILOT STUDY

ROSE SABAROFF
Oregon State University

The Problem

THE TEACHING OF reading with the aid of phonetic analysis has gone through several historical cycles: 1) early phonics (1890-1920) was mechanical and artificial, taught intensively, beginning with the first day in the first grade and constituting the entire reading program; 2) during 1920-35 or 40, the sterility of the older phonetic methods caused phonics to fall into disrepute, and it was largely eliminated from the program; 3) from 1940 to the present, the teaching of phonics has reached an all-time high in its second cycle. Unfortunately, however, for some teachers the resurgence of phonics has consisted of re-introducing the ancient mechanical system.

The great preponderance of research offers evidence to the effect that phonics is effective. The problem facing researchers today is how to teach phonics so that it is a more effective aid to reading.

Related Research

Several studies emphasize the value of a systematic phonics approach. Browne [1] (1939) found remedial readers in sixth grade were aided by a series of carefully planned lessons in phonics. Bedell and Nelson[2] (1954) found that when pupils of superior intelligence in grades 4, 5, and 6 were given a series of word-recognition lessons they made scores superior to a comparable group not given phonetic aid. House [3] (1941) found that fourth grade pupils who were given specific training with carefully prepared phonetic materials were distinctly superior to those who received no such training. Beltramo[4] (1954) used an alphabetical approach in the teaching of reading in grade 1 and concluded that for most children systematic instruction in phonics is important for gaining proficiency in these skills.

Studies by McDowell[5] (1935), Gates and Russell[6] (1938), Tate, Herbert, and Zeman[7] (1940) compared a conventional phonics approach with a functional one. They found that a moderate amount of phonics taught in a functional approach is superior to conventional phonetic drill. Two recent studies compared phonics taught from a series of special phonetic readers with a standard reading program and reached opposite conclusions. Sparks and Fay[8] (1957) found the group using Phonetic Keys to Learning series had an early advantage which was gone by fourth grade and concluded that children using the basic reading program gradually increase their phonetic skills as they progress through the grades and thus overtake those using the special series. Daniels and Diack[9] (1959), in a study in England, concluded that the systematic phonetic-controlled Royal Road Readers method showed itself decidedly superior to standard methods in developing power of word recognition for beginning readers.

The Present Study

The present study compares groups being taught by two phonetic methods and groups spending an equal amount of time in pleasure reading at their independent reading level. All groups continue to use the Basal Readers in a standard reading program in the classroom. The time spent with the experimenter is supplementary. The additional time spent in reading is equated for all groups using the three different methods. Gains made by low, average, and high achievers are compared. Differences among the methods are considered.

Research Design

	Low Achievers	Average Achievers	High Achievers
I	15 minutes Systematic Phonics	15 minutes Systematic Phonics	15 minutes Systematic Phonics
II	15 minutes Functional Phonics	15 minutes Functional Phonics	15 minutes Functional Phonics
III	15 minutes Independent Pleasure Reading	15 minutes Independent Pleasure Reading	15 minutes Independent Pleasure Reading

Population

The pupils taking part in the experiment were second graders from all three second-grade classrooms at Franklin Elementary School in Corvallis, Oregon. The pupils were grouped according to test results as low achievers (reading below grade level), average achievers (reading at grade level), and high achievers (reading above grade level). Within each of these ability groupings, pupils were then randomly assigned to a method of instruction. (Six per group; 18 per method; 54 in all).

Methods Used

The author taught all 9 experimental groups. She met at Franklin School with each group of six pupils, three times a week, 15 minutes per session, for 12 weeks during February, March, April, and May of 1960, for a total of 36 meetings (9 hours) per group.

Method I. Systematic Phonics (Taught separately to each of three ability groups.) With this method, the experimenter presented at each session those words and that reading matter necessary to acquaint the pupils with a given phonetic element. The sequence used was an adaptation of that of William H. Burton* and associates which is based on ease of learning and frequency of need in beginning reading. [10] All phonetic elements were presented in words and applied in meaningful reading matter immediately. Each ability group progressed through the same sequence, but at a different rate. For example, the slow achievers required three sessions with words using the short a sound and needed frequent review of that element while the high achievers sometimes could handle two or three phonetic elements at one session. The sequence used was as follows:

1. Consonants and digraphs. Auditory and visual perception of initial consonants in which errors were made on the preliminary phonics test. (The errors made and hence the consonants to be taught varied with each of the six phonics groups.) A consonant was taught in the initial position and then was used in words in the final and medial position also. For example, b in the initial position in ball, bump, bike; in the final position in cab, tub; in the medial position in baby, robin. The letter s was also used in its final position in nouns and verbs. The digraphs ch, th, wh, sh were treated the same way as single consonants.

2. Short a sound and short e sound. The short a sound was introduced before practice in all consonants had been completed. Thereafter, when a consonant or digraph was the new element being stressed, it was presented in words having the short a sound. The short e sound was presented next and handled in conjunction with consonants and digraphs in the same way. This was followed by mixed drill with words using known consonants and digraphs and the short a and short e sounds. Then additional consonants and digraphs as needed were taught using as examples words which contained the short a or short e sound.

3. Word endings: s, ed, ing; er, est; ly, y. These were added to known words or words that the pupils could now attack successfully with the phonetic elements they knew. Syllables were always underlined in words of more than one syllable.

4. Blends in which errors were made in the preliminary phonics test. (The blends needed varied in number and kind with each of the six phonics groups.) Bl, br, cl, dr, fr, gr, pl, pr, st, tr, fl,

*Formerly Director of Teacher Training at Harvard University, lecturer, and author of many books including Reading in Child Development.

sl, gl, sm, sn, sp, spr, st, str, sw, thr, tw, qu, and squ were taught in that sequence but at a different rate in each ability group and were interspersed with additional vowel work.

5. Short i, o, u sounds: ee, ea, ai, ay, oa. These vowel digraphs were often grasped more easily than the short i, o, and u sounds. These vowel digraphs were taught much as the consonant digraphs had been. The children were asked to respond to the two letters as one sight-sound element (ai together has the long a sound; ea together has the long e sound), rather than by the rule, "When two vowels come together, the first one has its long sound and the second one is silent." It seemed to the investigator that the pupils were able to use these sight-sound elements in attacking new words whereas the rule was memorized but not applied in reading. Also the rule is invalid in dipthongs and other vowel combinations.

6. Long vowel sound when a silent e appears at the end of a word, as in rid, ride; hat, hate; rod, rode. The pupils were led to generalize that it usually takes two vowels working together to make the long vowel sound.

7. Derivative words formed by adding a, un, ful, ish.

8. Recognizing syllables in words (not by rule citing).

9. Ar, er, ir, or, ur. These also were taught as single sight-sound elements. It was pointed out however, that r often affects the vowel proceding it in a word and that one should be aware of this.

10. Common dipthongs: oo, ou, ow, ew, oi, oy. The following materials were utilized in the presentation of the various phonetic elements and for interesting drill:

Building Reading Skills, Books 1, 2, and 3.[11]
Phono-Word Wheels[12]
Phonetic Word Drill Cards, Sets A, B, C.[13]
Take, A Sound Matching Game[14]
Phonics Flash Cards[15]
Stand Up - Sound Off![16]
Phonic Rummy, Sets A, B, C.
Group Sounding Game[18]
Webster Word Wheels[19]
The Syllable Game[20]

These materials were used selectively for drill on the particular phonetic element or elements being studied.

Method II. Functional Phonics (Taught separately to each of three ability groups). With this method the experimenter helped the pupils with the phonetic elements in words that were giving them difficulty in their immediate reading. When a pupil came to a word in his reading that he could not pronounce, the experimenter would put that word on the board and point out the phonetic element that would help the pupil sound the word. For example, if the word were train, the experimenter might write on the board words that start with the tr blend that the child already knew like tree, trip, etc. and also point out that ai in train makes the long a sound. The experimenter would then list other words with that same phonetic element — such as rain, paid, etc. Thus the phonetic elements of the particular word in reading were pointed out and also other known words that used the same element or elements.

With this method also, the experimenter sometimes drew out in advance the new words that were to appear in a story and helped the group attack the new words by pointing out the phonetic elements as described above.

The three groups using this method read from the Winston Series, a series they had not previously used in their classrooms. Each group of six pupils read the book at their instructional level as determined by preliminary testing. The low achievers read Good Stories, the high First Reader, pages 1-80; the average achievers read Along the Way, the low Second Reader, pages 1-102; the high achievers read Enchanting Stories, the high Third Reader, pages 1-105. Over the three-month experimental period, essentially the same phonetic elements were discussed as had been discussed in the systematic phonics groups except that the sequence of the elements to be analyzed was determined by the present need of the pupil in his reading rather than by a controlled sequential presentation. Thus, many more phonetic elements were discussed with a group during any given 15 minute period than was the case with systematic method.

Method III. Independent Pleasure Reading. These groups spent an equal amount of time in reading material of their own choice at their independent reading levels as determined by preliminary testing. These groups met in the library, and the experimenter made available to them a changing variety of books at the independent reading level of the group. She made sure that each pupil was reading a book that he could read with ease and enjoyment. These groups acted as controls since no phonetic aid was given. When a pupil needed help with a word, he raised his hand and the experimenter told him the word. Each pupil kept a record of the number and the titles of the books he read and recorded the words for which help was requested.

Psychological Principles

A survey of research indicates that there is common agreement on certain psychological principles of learning that should be observed in the proper teaching of phonics:

1. A mental age of 6.5 or 7 is considered desirable for successful attainment of phonetic skills.
2. Phonetic analysis should be related to needs-to reading a word not recognized in printed form but whose meaning is known.
3. Phonetic elements should be given direct

application in meaningful material.

4. A basic store of sight words is necessary.

Both of the phonetic methods employed in this study were in keeping with these principles.

Objectives of the Study

This was a pilot study to see if such a study can determine:

1. The comparative value of teaching phonics systematically versus teaching phonics functionally.

2. Whether the value of one or the other method varies with low, average, and high achievers.

It was hoped that light might also be shed incidentally on other problems of concern in the teaching of reading.

Data Gathered

Informal Reading Inventory

All the pupils were tested individually in a functional test of reading level by having them read 100-word selections from a graded Basal Reading Series and having them answer a few comprehension questions. The Houghton-Mifflin Series was used. This series had not been previously available to the pupils. If more than five significant errors were made per 100-word selection at grade level, then the child was given a selection to read in the book of the series that was one-half grade level lower, and so on to simpler and simpler material until the child could read with ease and understanding. This was then considered his instructional reading level. As for the pupil who read at grade level with almost no errors, the experimenter gave him increasingly difficult material to read until he reached the highest level in which he still made five or less significant errors and read with understanding. This then was his instructional reading level. This data was used to determine the grouping. Those pupils who read at grade level were grouped as average achievers. Those who read below grade level were considered low achievers; those who read above grade level, high achievers.

Standardized Test of Silent Reading Skills

The pupils were also given the Gates Advanced Primary Reading Test: Word Recognition, Form 1 and Paragraph Reading, Form 1.[21] The scores made on this test varied from half a grade higher to two grades higher than the grade level as determined by the informal reading inventory described above; 80% of the pupils measured a grade or more higher on the Gates test. However, this did not affect the grouping seriously. When the scores on the Gates test were arranged in rank order from highest to lowest, essentially the same children were in the top, middle, and low group. Where a discrepancy did suggest a shift in grouping, the experimenter retested the child individually. Thus the two tests acted as a check on one another. The informal reading inventory was considered to be nearer the child's true reading level for purposes of grouping for reading instruction. Form 2 of the Gates tests of word-recognition and paragraph reading was used for retesting at the conclusion of the experiment.

Basic Sight Word Test

The pupils were also given a basic sight word test based on Dolch's 220 basic service words.[22]

Phonics Test

The McKee Inventory of Phonetics Skill, Test One and Test Two was administered.[23] Knowledge of beginning consonants, ending consonants, digraphs, blends, short and long vowels, vowel combinations, inflected and derived word forms were tested.

Statistical Analysis

The experiment involved 54 children who were divided into three groups according to their reading ability. From each of these ability groupings, pupils were then randomly assigned into three subgroups to receive instruction by three different methods. This was, therefore, a completely randomized 3×3 factorial experiment with six replications.

For each child, four scores were obtained: sight words, phonetic elements, grade-level of word recognition, and grade level in paragraph reading. Gain in score during the experimental period was used as the variable. In addition, the average gain for word recognition and paragraph reading was used as the fifth variable.

The analysis of variance for these five variables is shown in Tables I to V. Table I is an Analysis of Variance for Gain in Sight Words. The three different ability groups of pupils show significant differences in gain in sight words. Table II is an analysis of variance for gain in phonetic elements. Pupils taught by the phonetic methods showed a significant gain in number of phonetic elements learned as compared with those who were not given this additional aid. The difference in method is significant at the 5% level. In addition, the difference in number of phonetic elements learned by pupils of the three ability levels is significant at the 1% level.

Table III shows Analysis of Variance for Gain in Word Recognition. Again, among the three ability groupings, there is a difference in gain, significant at the 1% level.

Table IV is Analysis of Variance for Gain in

TABLE I

ANALYSIS OF VARIANCE FOR GAIN IN SIGHT WORDS

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Method	25.48	2	12.74	1.36
Control vs. Expt.	25.04	1	25.04	2.69
Between Expt. Method	0.44	1	0.44	0.05
Level	145.37	2	72.69	7.80**
Method × Level	50.85	4	12.71	1.36
Error	419.33	45	9.32	
Total	641.04	53		

**Significant at 1% level.

TABLE II

ANALYSIS OF VARIANCE FOR GAIN IN PHONETIC ELEMENTS

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Method	248.04	2	124.02	3.95*
Control vs. Expt.	219.59	1	219.59	6.99*
Between Expt. Method	28.44	1	28.44	0.91
Level	499.26	2	249.63	7.95**
Method × Level	68.41	4	17.10	0.54
Error	1413.33	45	31.41	
Total	2229.04	53		

*Significant at 5% level.
**Significant at 1% level.

TABLE III

ANALYSIS OF VARIANCE FOR GAIN IN WORD RECOGNITION

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Method	0.0576	2	0.0288	0.40
Control vs. Expt.	0.0529	1	0.0529	0.73
Between Expt. Method	0.0047	1	0.0047	0.06
Level	2.6546	2	1.3273	18.23**
Method × Level	0.3689	4	0.1845	2.53
Error	3.2767	45	0.0728	
Total	6.3578	53		

** Significant at 1% level.

TABLE IV

ANALYSIS OF VARIANCE FOR GAIN IN PARAGRAPH READING

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Method	0.5867	2	0.2934	1.10
Control vs. Expt.	0.5735	1	0.5735	2.15
Between Expt. Methods	0.0132	1	0.0132	0.05
Level	1.2420	2	0.6210	2.32**
Method × Level	0.7840	4	0.1960	0.73
Error	12.0236	45	0.2672	
Total	14.2763	53		

**Significant at 1% level.

TABLE V

ANALYSIS OF VARIANCE FOR GAIN IN AVERAGE OF WORD RECOGNITION AND PARAGRAPH READING

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Method	0.25208	2	0.12604	1.63
Control vs. Expt.	0.24368	1	0.24368	3.15
Between Expt. Methods	0.00840	1	0.00840	0.11
Level	1.75278	2	0.87639	11.32**
Method × Level	0.31121	4	0.07780	1.00
Error	3.48387	45	0.07742	
Total	5.79994	53		

**Significant at 1% level.

TABLE VI

MEAN GAIN FOR THREE ABILITY GROUPS

Variable	Low	Average	High
Sight	3.72	0.39	0.11
Phonics	9.39	4.83	2.00
Word Recognition	0.23	0.69	0.20
Paragraph Reading	0.37	0.74	0.54
Av. of Word and Para.	0.30	0.71	0.37

TABLE VII

MEAN GAIN FOR THREE METHODS OF TEACHING

Variable	Independent	Systematic	Functional
Sight	0.44	2.00	1.78
Phonics	2.56	7.72	5.94
Word Recognition	0.33	0.38	0.41
Paragraph Reading	0.40	0.60	0.64
Av. of Word and Para.	0.37	0.49	0.52

Paragraph Reading. There is no significant difference among the groups in this variable. However, in Table V where the gain in word recognition and paragraph reading are combined, there is a difference in gains made by the three ability groupings, significant at the 1% level.

The unexpected result was that the significant differences in learning were related in the main to the level of reading ability with which the pupils began rather than with the method used. The only area in which the difference in method showed significantly was in the number of phonetic elements the pupils learned.

Table VI summarizes the mean gain for the three ability groupings. As can be clearly seen, the low achievers made the largest gain in sight words and in phonetic elements learned. In word recognition, paragraph reading, and combined word recognition and paragraph reading, the average group made the greatest gains. These latter gains are gains in grade-level of reading and show a mean gain of .71 grades for the average readers during a 12-week period, after a total of only nine hours of reading in addition to regular classroom instruction. The gains made by the low and high achievers in these areas are not nearly so outstanding.

Table VII summarizes the mean gain for the three methods of teaching. The only significant difference among the methods of instruction is in the area of phonics. There is a significant difference between the number of phonetic elements learned by the pupils taught by the two phonetic methods and that of the pupils doing independent pleasure reading. The difference in phonetic elements learned by means of the two different phonetic methods is not significant, however.

Conclusions and Projections for Future Study

The major objectives of this pilot study were to see if a study could be devised that would determine: 1) the comparative value of teaching phonics systematically versus teaching phonics functionally and 2) whether one or the other method was more effective with low, average, or high achievers in reading.

The results of the present study were suggestive, not conclusive, in relation to these major objectives. As we see in Table VII, more phonetic elements were mastered by the systematic method than by the functional method. The average gains in paragraph reading in the three-month period were .64 grades with the functional phonics method, .60 grades with the systematic phonics method, and .40 grades with independent pleasure reading. Those groups taught by the phonetic methods learned significantly more phonetic elements than the independent reading groups. If the experimental period had been longer, the possible superiority of one or other of the phonetic methods might have been established Also, with a longer instructional period, the increased knowledge of phonics might have become more evident in the word-recognition and paragraph-reading scores. Since the low achievers made by far the largest gains in phonetic elements (Table VI), a carry-over to improved word recognition and paragraph reading is much to be desired. In a future study it might be very worthwhile to evaluate the transfer phonetic knowledge to reading skills.

In order to achieve the objectives set forth above, this pilot study seems to suggest the following conditions for a future study:

1. The period of instruction should be longer, at least a year, preferably two.
2. The experimental method should completely replace the classroom teaching instead of being supplemental to classroom teaching. The differences in effectiveness possible between the methods may well have been negated by the fact that all the children were getting good basal instruction in reading in their own classrooms which included teaching of phonics even for those who read for pleasure at their independent reading level and received no additional phonetic instruction during the experimental period. The mere addition of nine hours of phonetic instruction by two different methods would not seem to be sufficient to produce a significant difference in reading ability among the groups.
3. The sample should be larger. There should be more pupils in each subgroup because of the large variation appearing among the pupils.
4. The testing procedures should be improved. A more refined test of word recognition and paragraph reading appears to be needed. One or two errors or correct guesses caused too great a shift in total test score on the silent reading test used. Also tests with higher ceilings should be selected to permit the high achievers to show their true gains. It might be worthwhile to have informal reading inventories given at the conclusion of the experimental period as well as at its commencement as a further check.

Other Implications

Do all pupils benefit equally from a good program of phonics?

According to the agreed upon psychological principles set forth in this paper, both programs being tested in this study were "good" programs. Yet it is obvious from the scores that pupils' gains varied greatly within any group. This study has demonstrated emphatically that individual differences are very important! The most significant differences statistically were those among low, average, and high achievers.

Who benefited most?

It would seem from Table VI that the low achievers made the greatest gains in phonetic elements, but the average achievers made the greatest gains

in word recognition and paragraph reading.

<u>Do certain undesirable concomitants arise in either of the phonetic methods?</u>

While working with the pupils in the phonics groups, the experimenter soon became aware that became aware that some of the pupils were unable to use phonics effectively in attacking a word. They would sound an initial consonant and get no further in the word, or they might sound the individual letters of a short word yet not grasp the word as a whole. These children made poor gains in reading despite much effort on their part as well as that of the experimenter. These children possibly needed an entirely different means of word attack, such as the VAKT (a combined sounding-as-you-trace technique) which this experiment was not set up to provide.

<u>How do scores on a phonics test correlate to scores in a silent reading test of word recognition and paragraph reading?</u>

The correlation coefficient* between phonetic elements known and the scores on Gates Tests of Word Recognition and Paragraph Reading was .66 at the conclusion of the experimental period. At the end of the three months, those pupils whose paragraph reading was between 2nd and 3rd grade level scored an average of 68.6 phonetic elements. Those pupils reading between 3rd and 4th grade level showed knowledge of 76.8 phonetic elements. Those reading at 4th grade level or above scored 81.4 phonetic elements. The highest score possible was 84 phonetic elements. Though we see that the better readers had greater mastery of phonetic elements, we cannot impute a cause-effect relationship. In further analyzing the scores, the experimenter found that 25 of the 28 pupils who made <u>gains</u> of .7 grades or more during the experimental period had mastered 76 to 84 phonetic elements, with the average at 80. Again, knowledge of phonics was evident. Yet, not all pupils with scores of 76 phonetic elements or above achieved an increase of .7 grades which re-emphasizes the fact that knowledge of phonics is not the whole reading story.

<u>How does grade level of reading as determined by an informal reading inventory given individually compare with the score a standardized group test of reading?</u>

The scores on the standardized test were higher in all cases than the child's functional ability to read selections in graded reading material. With these second-graders, the scores on the standardized test varied from half a grade to two grades higher. Eighty percent of the children measured a grade or more higher. From this we might conclude that for determining the level of reading at which a given child functions in the classroom, the informal reading inventory is definitely more dependable.

<u>Does easy pleasure reading contribute to the development of reading skills?</u>

The present experiment was not set up to test the effectiveness of easy pleasure reading in development of reading skills. Yet some of the data derived may merit perusal. Eight of the eighteen pupils in the groups reading for pleasure at their independent reading level gained .8 grades or more in the three-month experimental period. The low achievers averaged the least gain in paragraph reading of the groups using this method. (The low achievers gained .21 grades during the three-month period as against .51 grades for the average achievers and .48 grades for the high achievers). The average gain for the three groups using this method was .40 grades compared to .60 grades with systematic phonics and .64 grades with functional phonics.

Summary

The main objectives of this pilot study were 1) to investigate the comparative value of two methods of teaching phonics (systematic and functional) with a control group reading for pleasure at the independent reading level of the pupils and 2) to determine whether one or the other method was more effective with low, average, or high achievers. The additional time spent in reading was equated for all groups.

The results obtained during the three-month experimental period were inconclusive but suggest that more phonetic elements were mastered by the systematic method (Table VII). Analysis of the raw scores** suggest that phonics taught systematically may be superior for low achievers and functional phonics superior for average and high achievers. In order to verify these suggestive outcomes, a future study should: 1) extend the period of instruction to at least a year, preferably two, 2) use the experimental method in the classroom rather than as a supplement to classroom teaching, 3) increase the size of the sample because of the large variation appearing among pupils in each subgroup, and 4 improve the testing. The tests of word recognition and paragraph reading will have to be more refined and have a higher ceiling.

The most outstanding result in the pilot study was the emphasis given to individual differences in learning ability. The most significant differences in learning were between the ability groups rather than between the methods used. Since the investigator taught all the groups, the quality of teaching should have been essentially the same, yet the average achievers made by far the largest gains in word recognition and paragraph reading regardless

* Karl Pearson's product-moment correlation coefficient.

** Raw scores are available on request.

of method. This should not be taken to mean that any method, good or bad, is acceptable since all the methods used in this study were psychologically "good" methods. Neither does it mean that the high and average achievers will make the same strides no matter what the classroom teacher does and that low achievers will continue to make poor gains or no gains. But it does show that no matter how hard the teacher and children try, all children cannot be expected to make the same amount of gain.

Those who made higher scores in reading knew more phonics on the average. However, all children who scored high in phonics did not read equally well. It seemed from the analysis of scores that teaching of phonetic elements will not solve the whole reading problem. Not all pupils grasped the phonetic elements to which they were exposed, and not all pupils showed an ability to read in keeping with the phonics they knew.

It is hoped that a larger and more extended study will follow to substantiate the results suggested by the present pilot study.

REFERENCES

1. Browne, M. Dorothy. Phonics as a Basis for Improvement in Reading. Washington, D.C., the Catholic University of America, 1939.
2. Bedell, Ralph and Nelson, Eloise S., "Word Attack as a Factor in Reading Achievement in the Elementary School," Educational and Psychological Measurement, V. 14 (Spring, 1954), 168-175.
3. House, Ralph W. "The Effect of a Program of Initial Instruction on the Pronunciation Skills at 4th Grade Level as Evidence in Skills Growth," Journal of Experimental Psychology, V. 10 (Sept., 1941), 54-55.
4. Beltramo, Louise. "An Alphabetical Approach to the Teaching of Reading in Grade One," Dissertation Abstracts, V. 14, Part 3 (1954), 2290.
5. McDowell, John B. "A Report on the Phonetic Method of Teaching Children to Read," Catholic Educational Review, V. 51 (Oct., 1953), 506-519.
6. Gates, Arthur and Russell, David H. "Types of Material, Vocabulary Burden, Word Analysis, and other Factors in Beginning Reading," Elementary School Journal, V. 39 (Sept. and Oct., 1938), 27-35, 119-128.
7. Tate, Harry L., Herbert, Theresa, and Zeman, Josephine K. "Non-Phonic Primary Reading," Elementary School Journal, V. 40 (Mar., 1940), 529-537.
8. Sparks, Paul E., and Fay, Leo C. "An Evolution of Two Methods of Teaching Reading," Elementary School Journal, V. 56 (Apr. 1957), 386-390.
9. Daniels, J.C. and Diack, Hunter. "The Phonetic Word Method," The Reading Teacher, (Oct., 1959), 14-21.
10. Burton, William H.; Baker, Clara Belle; Kemp, Grace K. "A Complete Basic Program for Grades One Through Three," Reading for Living Series, 18-21. A Bobbs-Merrill Company Booklet.
11. Building Reading Skills, McCormick-Mathers Publishing Co.
12. Phono-Word Wheels, Steck Co., Publishers.
13. Phonetic Word Drill Cards, Sets A, B, C. Kenworthy Educational Services, Inc.
14. Take, A Sound Matching Game, Gerrard Press.
15. Phonics Flash Cards, Gelles-Widmer Co.
16. Stand Up – Sound Off! Charles E. Merrill Books.
17. Phonic Rummy, Sets A, B, C. Kenworthy Educational Services, Inc.
18. Group Sounding Game. Gerrard Press.
19. Webster Word Wheels, Webster Publishing Co.
20. The Syllable Game. Gerrard Press.
21. Gates Advanced Primary Reading Test, Bureau of Publications, Teachers College, Columbia University.
22. Basic Sight Word Test. Gerrard Press.
23. McKee Inventory of Phonetic Skill, Test One and Test Two, Houghton-Mifflin Co.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

GROUP INFLUENCE ON CREATIVITY IN MATHEMATICS[1]*

FRANK W. BANGHART
University of Virginia
HAROLD S. SPRAKER
Middle Tennessee State College

THE WORD creativity has various connotations. Hadamard (7) and Welch (22) thought of creativity as combining or recombining ideas in conformity with some plan. Crawford (4) states that creation is not merely a matter of combination but also consists of the shifting of attributes from one thing to another. Stein refers to creative work as that which is novel, useful, or perhaps satisfying to a group. Guilford (5) suggests that creativity is not a unity, but is rather a collection of different component abilities or traits. Wilson, Guilford and Christianson (24) state that one of the most important aspects of creativity is originality. The variations of the concept of creativity are suggested by Morgan (13) who has compiled 25 definitions of creativity, and Rhodes (16) who has extended the definitions to 32. Rhodes' definitions of creativity might be grouped into eight broad categories. These are: mysticism, potentiality and ability, process, adaptation, problem solving, synthesis, pattern of behavior, and originality.

It is suggested by several authors that inventors, artists, etc., are especially endowed with the creative gift. However, authors, such as Guilford, feel this is not necessarily true, that all persons are endowed with creative potential, the difference primarily being that some people manifest the creative abilities more than others. Rogers (17), Hill (8) and Osborne (14) also agree t at creative talent probably exists in all individuals. During the past decade, Guilford has conducted a series of factor analysis studies of creative thinking at the University of Southern California. These factors are enumerated in an article by Wilson, Guilford, Christianson and Lewis (23).

The headings are listed here as verbal comprehension, numerical facility, perceptual speed, visualization, general reasoning, sensitivity to problem, work fluency, associational fluency, ideational fluency, adaptive flexibility, spontaneous flexibility, originality, closure, redefinition, or judgment. Lowenfeld (10) has also reported a factor analytic study on creativity at Penn State.

In Lowenfeld's report, he has isolated eight criteria of creativity that he claims significantly differentiated between creative and less creative people. This work has been conducted independently of the work of Guilford. However, it is notable that there is a significant amount of similarity between the two studies. The common factors reported by both studies refer to the concept of sensitivity, fluency of ideas, flexibility, and originality. It's possible that some of the other discrepancies between the two studies are accountable by the more or less subjective assignment of names to the factors derived.

There does not seem to be an exceedingly high correlation between intelligence and creativity. In L. S. Hollingworth's study on gifted children (9), she says "If a general statement be attempted on the basis of such data as the description that is in the summaries, it might be to the effect that one third of these highly intelligent children show notable signs of creativeness. Another one third show such indications to a moderate degree. In the remaining one third, there is at least no indication or marked constructive originality provided by these descriptions."

Witty (25) has expressed also the same opinion: "If by gifted children, we mean those youngsters who give promise of creativity of the high order, it is doubtful if the typical intelligence test is suitable for use in identifying them. For creativity posits originality and originality implies successful management control and organization of new materials or experiences. The content of the intelligence tests is blatantly lacking in situations which disclose originality or creativity."

In 1950, Thurstone (21) also expressed a similar view when he said "To be extremely intelligent is not the same as to be gifted in creative work." This may be taken as an hypothesis. It is a common observation in the universities that those students who have high intelligence, judging by available criteria, are not necessarily the ones who produce the most original ideas. There is undoubtedly a positive correlation between creative talent and intelligence, so that geniuses are usually in the upper half of the general intelligence distribution. Guilford (6) has suggested that the low correlation between creativ-

* Footnotes will be found at the end of the article.

ity and intelligence is attributable to the fact that different intelligence tests do not intercorrelate perfectly even when errors of measurement have been taken into account,and that each test emphasizes a different pattern of primary abilities.

Mier and Stein (12) reported a study of 64 research chemists when correlations were made between creativity as determined by the department heads. The Wechster-Bellevue Intelligence Scale and the Miller Analogies Test were used. The correlation between creativity and the Miller Analogies was -.18 for Ph.D.'s. Also for Ph.D. chemists, the correlation between Wechsler-Bellevue and creativity was .46.

Mandell (11) reports in a study dealing with the correlation between originality and achievement with 35 physicists. He found a correlation of .59. It is interesting to note that the correlation coefficient for basic tests in chemistry given to 55 chemists was .46. From the study of the literature, it can be concluded that intelligence and creativity complement one another, but are not necessarily synonomous.

Undoubtedly the general behavioral patterns which the individual develops has some bearing upon his creative functions. Thomas (20) has stated that some attributes of the creative mind are accepted almost as axioms— curiosity, imagination, enthusiasm, and a high level of mental energy. Rosenfeld (18) suggests that curiosity is the beginning of creative learning. Barron (2) reports on a study of 100 officers where attempts were made to determine the relationship between originality and other aspects of personal and intellectual functioning.

The findings might be summarized under the headings a) a disposition toward integration of diverse stimuli; b) energy-fluid output and involvement; c) personal dominance and self-assertion; d) a responsiveness to impulse and emotion; e) an expressed femininity of interests; f) general effectiveness of performance. In another report, Barron (3) suggests that the disposition toward originality may thus be seen as a highly organized mode of responding to experience. The socially undesirable traits which may go along with it include rebelliousness, disorderliness, exhibitionism, while the socially valued traits which accompany it include independence of judgment, freedom of expression, and novelty of construction and insight.

Stein (19) reports that some investigators who have relied primarily on the Rorschach tests in studying creativity have remarked that the records of the subjects contain responses that are to be found in the records of severely disturbed persons. These responses are in sufficiently large numbers that the emotional stability of the creative person is questioned.

Thomas (20) makes the statement concerning the motivation to create, "We have stated that, based on subjective judgment, the motivating factors that influence men to dedicate these attributes to the life of creativity are in descending order of importance. The human ego's appetite for distinctive definition--the individual's personal urge to seek self-fulfillment in a unique manner of his own choosing. We have asserted that another incentive is rooted in a person's heightened consciousness of social approval and finally we have suggested that there may be a third source or stimulus to creativity much more questionable than the others--the desire to attain material success."

In terms of intrinsic motivational forces, a number of writers tend to emphasize the importance of the four conditions which follow: 1) intensive interest in something; 2) freedom to nourish and express the interest; 3) personal security which permits nonconformity; 4) a time for reflective thought."

Osborne (14) has pointed out the importance of environment upon the development of creativity. He writes "Most of us are highly imaginative in childhood and yet many of us grow up to be non-creative. One reason for this may be that as a nation we have not made enough of the importance of ideas. Another reason is that most parents are guilty either of active discouragement or at least of lack of active encouragement of the children.

In an address delivered to the Fourth Annual Creative Problem Solving Institute in June of 1958, Allen (1) reported on the search for some basic problems that were encountered in both industry and in the schools, which involved the creative process. The major problem identified was "the almost universal reluctance or inability to think creatively about problems." He found that from three to five years of additional training of carefully selected science and engineering graduates were necessary for them to be able to perform the work expected of them at the General Electric Company at Schenectady. The difficulty was not that of any lack of knowledge or the ability to use that knowledge in solving a problem, but the weakness lay rather in the dearth of habit and willingness to propose several possible solutions to the assigned problem. The implication here is that something can be done about nurturing creativity. In this regard, Guilford (6) has said "I will venture one or two opinions on a general problem of the development of creativity, for I believe that much can be done to encourage its development. This development might be made in the nature of actual strengthening of the functions involved, or it might mean the better utilization of what resources the individual possesses, or both."

Purpose of the Present Study

The present study was designed to investigate the role of group influence on creativity during mathematical problem solving. An attempt was made to isolate the effects of group influence and to determine whether or not the subjects benefited from active participation within a group. The hypothesis to

be tested concerned the possible effects of the group input upon creativity and the relative merits of the subjects creativity as a result of working with a group versus working alone.

Subjects

The subjects were selected from intact seventh grade classrooms located in an urban junior high school. In all, 180 subjects participated in the experiment. Eighty-seven subjects received the group treatment and 93 received the individual treatment. Fifty-four subjects were in the two high classes, 95 in the four average classes, and 31 in the two low classes.

Treatments

The experiment was conducted under two treatment conditions. Treatment I involved the subjects in the study of mathematics under conditions which permitted them to work as a group, interchanging information and helping one another. Treatment II imposed conditions whereby the subjects could not exchange information or help one another. Solutions of the problems were dependent upon the individuals' resources.

Materials

During the present experiment, Units II and III from Mathematics Units, Volume I, Number Systems, by the School Mathematics Study Group[2] were used. The teachers also used a commentary[3] for these units written by the same group. Unit II of this book deals with numeration. An attempt is made to deepen the pupil's understanding of decimal notation and to increase his understanding of the reasons for the algorithms of addition, subtraction, multiplication, and division. The unit first discusses some ancient systems of numeration. These systems are used as a contrast to the Hindu-Arabic system in which the symbol for a number and its position both have significance. The characteristics of some of these systems are demonstrated in order for the pupils to appreciate the important characteristics of their own system.

The unit deals next with numeral systems with a base other than ten. In using a new base, the students are forced to look at reasons for carrying, borrowing, and other mechanical operations. Through this, they should gain deeper insight into the decimal system of numeration.

The following[4] is an illustration of how the material is developed in this unit:

Add these numbers in base ten:

$$\begin{array}{r} 25 \\ \underline{48} \end{array}$$

As you know, you "carry" when the sum of a column is ten or more,

$$\begin{array}{rl} 25 = & 2 \text{ tens} + 5 \text{ ones} \\ \underline{48 =} & \underline{4 \text{ tens} + 8 \text{ ones}} \\ & 6 \text{ tens} + 13 \text{ ones} \\ = & 7 \text{ tens} + 3 \text{ ones} \end{array}$$

$$\begin{array}{rl} \text{Add: } 24 \text{ seven} = & 2 \text{ sevens} + 4 \text{ ones} \\ \underline{35 \text{ seven} =} & \underline{3 \text{ sevens} + 5 \text{ ones}} \\ & \underline{5 \text{ sevens} + 9 \text{ ones}} = 6 \text{ seven} + 2 \text{ ones} \\ & = 62 \text{ seven} \end{array}$$

Is it possible to "carry" in addition in base seven just as you do in base ten?

In base ten, you "carry" when the sum of a column is ? or more.

In base seven, you "carry" when the sum of a column is ? or more.

Unit III deals with the natural numbers and zero. An attempt is made in this unit to develop in students the realization that number is an abstract concept—an idea rather than something that can be seen or felt. The unit discusses at length the commutative, associative, and distributive laws of the arithmetic operations and the special characteristics of zero and one. The closure property and the idea of inverse operations are also discussed briefly.

An illustration of the development of some of the material in Unit III follows:

> The number zero is even more special than the number one. It can be used to count in the sense that if you have no apples, you can express that fact by saying that you have sero apples. It is very useful in our notation for numbers since it serves often as a place-holder...
>
> Can we divide by zero? We know that $6/2$ is 3 because $3 \times 2 = 6$. So if $3/0$ is a number, it should be one which when multiplied by zero gives three. All the numbers we have had so far give zero when multiplied by zero. It would be very strange to have a number such that when we multiply it by zero, we obtain three. Another difficulty would be that if $3/0$ were a number, it would be equal to 1.0 since we could divide numerator and denominator by 3. Then $3/0$ and $1/0$ would be equal and yet when you multiplied the first by zero, you would get 1. For these and other reasons, we exclude division by zero.[5]

These units were chosen for use in the experiment for four reasons. First, they have been carefully written and used experimentally in School Mathematics Study Group Centers throughout the United States. Seventy-five teachers taught Unit II in 110 classes; seventy-six teachers taught Unit III in a comparable number of classes during 1958-59. There is general agreement among those teachers that the units are useful.[6] Second, the units are designed to be used either as supplements to the regular course or as replacements for certain parts of the regular course for the seventh grade. Third, they contain useful mathematical ideas which are new to the pupils, thereby eliminating some of the effects of previous practice. Finally, these units

were developed in such a way that they promote creativity in learning mathematics.

Procedure

The procedure in the daily class was outlined as follows: the teacher briefly discussed the topic under consideration. Insofar as possible, the teacher did not explain a method of solution to the pupils, but encouraged them to find a method for themselves. The teacher also promoted creativity in other ways. This was done by 1) encouraging unique methods of solving problems, 2) asking for several methods of solving a single problem, and 3) encouraging independent generalization of these various methods of solution. Any method of solution which was correct was approved by the teachers regardless of the efficiency of the method.

In order to determine whether a systematic difference among the classes was introduced by the teachers, a t test was used between the creativity scores of the classes taught by the two teachers. The t test was not significant. This suggested that the slight difference in creativity scores was due to chance rather than any superiority of either of the teachers.

Measurement Indices

The pupils first took a standardized arithmetic achievement test and a mental maturity test. They next received instruction on Unit II. Then, the first part of the creativity test was administered. The pupils then received instruction on Unit III and took the last part of the creativity test. A creativity test was administered in two parts, the first part being given after the students had studied Unit II for three weeks and the second part at the end of the experiment. Two days of testing were needed for the administration of each part of the tests. Some items measuring creativity were also given to be done as homework. They were allowed no help with those items dome as homework, but the time used to work on this part of the test was not limited.

The test designed to measure mathematical creativity was composed of problems in mathematics with instructions to the subjects to solve each problem in as many different ways as they could. This was the type of item used by Guilford to measure the originality factor. It was also felt that this test would measure the factors of fluency and flexibility; Guilford suggests that creative performance is distinguished by its dependence on these two factors and originality. [7] Fluency, the ability to produce a relatively large number of ideas per unit of time, would obviously have a bearing on the number of different methods of solution to a problem. Flexibility, the ability to adapt to changing instructions or situations, would be involved in the shift from one type of problem to the next.

The test of mathematical creativity was composed of mathematics problems with instructions to solve each in as many ways as possible. When the subjects had completed the test, a statistical check was made of the frequencies with which each type of response was made. This was done separately for each problem. A response was then weighted according to its frequency of occurrence. Responses given by one fourth or less of the subjects were assigned weights of four; responses given by one-fourth to one-half of the subjects were assigned weights of two, and responses given by three-fourths or more of the subjects were assigned weights of one.

The division of the responses and the assignment of the weights are similar to those used in Guilford's studies. It was determined that this range of weights was sufficiently large to give adequate dispersion in total scores and sufficiently small for reasonably convenient scoring. Each subject's responses were assigned the appropriate weights and the sum of the weights for all responses was the creativity score. The split-half reliability of the test was .82.

Analysis of Data

The analysis of data was accomplished by means of an analysis of covariance design, the t test, and two correlational computations. The analysis of covariance was based on group and individual treatments, adjusting for possible differences in IQ and arithmetic achievement. The t test was used to check possible differences among the high, average, and low classes re creativity. The correlations were used to indicate statistically the relationship between IQ and creativity and between arithmetic achievement and creativity.

Table I shows the mean scores by treatment in creativity, California IQ, and California Arithmetic Achievement. It is apparent from this table that those pupils working as individuals have a higher creativity score than those working in groups. However, it can also be seen that the pupils serving as individuals have a slightly higher IQ and arithmetic achievement score than those working in groups. The question to be answered is whether the difference in the creativity scores for the two treatments is significant when adjustments are made for differences in IQ and arithmetic achievement. Analysis of covariance was used to check the difference.

Analysis of covariance was used to control the concomitant variables, achievement and intelligence. Table II summarizes the covariance computations. It is apparent that no significant difference exists between group and individual problem solvers when adjustments have been made for intelligence and achievement. The results are consistent with previous studies which suggest that the contribution of group activity is illusory in many if not most problem solving situations.

TABLE I

MEAN SCORES BY GROUP AND INDIVIDUAL TREATMENTS

Treatment	Creativity	California Mental Maturity--IQ	California Arithmetic Achievement
Group	103.99	104.47	74.83
Individual	111.92	106.98	82.18

TABLE II

ANALYSIS OF COVARIANCE

Source	cj	ss	ms	f
Total	177	124,413.04		
Within	1	124,340.95	706.48	.10
Between	176	72.09	72.09	

TABLE III

MEAN SCORES BY LEVEL

Level	Creativity	California Mental Maturity--IQ	California Arithmetic Achievement
High	142.481	118.037	96.815
Average	99.242	103.526	74.821
Low	75.290	91.258	58.612

TABLE IV

MEANS, STANDARD DEVIATIONS AND t TESTS AMONG THE THREE LEVELS

Statistic	High	Average	Low
M	142.481	99.242	75.290
s	34.064	27.223	20.209
N	54	95	31

tHA = 7.989* tHL = 11.413* tAL = 5,230*

Key: tHA = t test between high and average classes.
tHL = $\bar{t}$ test between high and low classes.
tAL = $\bar{t}$ test between average and low classes.
* Significant beyond the .01 level.

Since there was no significant difference in the creativity scores of pupils studying as individuals and pupils studying in groups, these scores were combined and the t test was used to test for the significance of differences among the high, average, and low classes involved in the experiment. Pupils were assigned to these classes by school officials on the basis of IQ, achievement, and the recommendations of the teachers.

Table V shows the mean creativity, IQ, and arithmetic achievement scores of the high, average, and low classes. It is apparent from Table III that the low classes have the smallest creativity score, the average classes a higher creativity score, and the high classes the highest creativity score. The question to be answered by the t tests is whether the differences in the creativity scores at the different levels are significant.

Tab IV contains means, the standard error of the means, N, and the t ratio. As shown in Table IV, each of the t tests is significant beyond the .01 level. It is concluded that the high classes are more creative than the average and low classes and that the average classes are more creative than the low classes.

Creativity

In addition to the analysis of covariance and the t test, product moment correlation coefficients were found between creativity and IQ and between creativity and arithmetic achievement. The correlation coefficients between this creativity measure and IQ (.59) and between creativity and arithmetic achievement (.66) suggest a moderate positive relationship between creativity and each of these variables.

These coefficients are in general agreement with those found by other investigators. These other investigators also found a moderate positive correlation between creativity and IQ and achievement.

Summary

The results of this investigation suggest that no significant difference exists between group and individual problem solvers in terms of creativity. However, significant differences do exist between different levels of intelligence. A derived correlation coefficient of .59 was found between creativity and intelligence, and .66 between creativity and achievement.

This report culminates a series of studies on group problem solving conducted at the University of Virginia under the sponsorship of the Group Psychology Branch of the Office of Naval Research. As a result of five separate pieces of research dealing with the relative merits of group problem solving it is concluded that the contribution of the group has been overly emphasized. In none of the five research studies completed did the group factor make any contribution to problem solving. On the contrary, there seems to be a consistent, if slight, advantage to solving problems alone.

FOOTNOTES

1. This study was made possible by a research grant from the Group Psychology Branch, Office of Naval Research, Contract No. Nonv 474 (8).
2. School Mathematics Study Group, Mathematics Units, Volume I, Number Systems (New Haven, Connecticut: School Mathematics Study Group, 1959), pp. 1-54.
3. School Mathematics Study Group, Junior High School Mathematics Units, Volume I, Commentary for Teachers (New Haven, Connecticut: School Mathematics Study Group, 1959), pp. 1-63.
4. School Mathematics Study Group, op. cit., p.11.
5. Ibid., pp. 52-53.
6. School Mathematics Study Group, Teachers' Reports on Experimental Units for Grades 7 and 8, Newsletter No. 3, (New Haven, Connecticut: School Mathematics Study Group, September, 1959), pp. 5-7.
7. IIj. P. Guilford, A Revised Structure of Intellect, Report from the Psychological Laboratory No. 19 (Los Angeles: University of Southern California, 1957), 19.

REFERENCES

1. Allen, M. S. Technologies Need for More Creativity in Education (New York: Creative Education Foundation, 1958), pp. 1-11.
2. Barron, F. "Originality in Relation to Personality and Intellect," Journal of Personality, XXV (December 1957), pp. 730-42.
3. Barron, F. "The Disposition Toward Originality," Journal of Abnormal and Social Psychology, LI (November 1955) pp. 478-85.
4. Crawford, Robert P. The Techniques of Creative Thinking (New York: Hawthorn Books, Inc. 1954), p. 50.
5. Guilford, J. P. "Creative Abilities in the Arts," Psychological Review, LXIV (March 1957), p. 110.
6. Guilford, J. P. "Creativity," The American Psychologist, V (September 1950), p. 448.
7. Hadamard, Jacques. An Essay on the Psychology of Invention in the Mathematical Field (New York: Dover Publications, Inc.), 1954 p. 104.
8. Hill, W., et al. How Children Can Be Creative, Office of Education, United States Department of Health, Education and Welfare, Bulletin No. 12 (Washington: Government Printing Office, 1954), p. 1.
9. Hollingsworth, L. S. Gifted Children: Their Nature and Nurture (New York: Macmillan, 1926).

10. Lowenfeld, Viktor. "Current Research on Creativity," National Education Association Journal, XLVII (November 1958),pp. 538-40.
11. Mandell, M. M. "Measuring Originality in the Physical Sciences," Educational and Psychological Measurement, X (Autumn 1950), pp. 380 -385.
12. Meir, B., and Stein, N. I. "Measures of Intelligence and Creativity," Journal of Psychology, XXXIX (January 1955), pp. 117-26.
13. Morgan, D. N. "Creativity Today," Journal of Aesthetics and Art Criticism, XII (January 1953), pp. 1-24.
14. Osborne, A. G. Applied Imagination (New York: Charles Scribners and Sons, 1953).
15. Osborne, op. cit., pp. 26-41.
16. Rhodes, J. M. "The Dynamics of Creativity," (Unpublished Dictors Dissertation, University of Arizona, 1956), pp. 13-18.
17. Rogers, C. R. "Toward a Theory of Creativity," ETC., XI (May 1954), pp. 249-60.
18. Rosenfeld, P. "Creative Learning," Educational Forum, XXII (May 1958).
19. Stein, M. I. "Creativity and Culture," Journal of Psychology, XXXVI (October 1953), p. 315.
20. Thomas, C. A. "Creativity in Science," The Eighth Annual Arthur Dehon Little Memorial Lecture (Cambridge, Massachusetts Institute of Technology, 1955), p. 9.
21. Thurstone, L. L. Creative Talent, Psychometric Laboratory, University of Chicago, LXI (December 1950).
22. Welch, Livingston. "Recombination of Ideas in Creative Thinking," Journal of Applied Psychology, XXX (December 1946), p. 642.
23. Wilson, R. C., Guilford, J. P., Christensen, P. R., and Lewis, D. J. "A Factor- Analytic Study of Creative Thinking Abilities," Psychometrika, XIX (December 1954), pp. 291-311).
24. Wilson, R. C., Guilford, J. P., and Christensen, P. R. "The Measurement of Individual Differences in Originality," Psychological Bulletin, L (September 1953), p. 362.
25. Witty, P. A. "Contributions to the IQ Controversy from the Study of Superior Deviates," School and Society, II (April 1940), p. 504.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

A COMPARATIVE ANALYSIS OF PUBLIC SCHOOL FINANCE, PERSONNEL AND PUPILS

B. EVERARD BLANCHARD
DePaul University

ONE OF THE many criticisms directed at the public school system these days is of the vast amount of money being spent to maintain the schools according to the dictates of society. If we continue to spend more and more money, can we reasonably expect to receive better educational returns? Research studies have tended to support the conclusion that pupils attending high-expenditure elementary and secondary schools, on the average, achieve higher scores in the academic subjects than do pupils attending low-expenditure schools. In addition, high-expenditure school districts have the money to employ teachers with better preparation than one would find in low-expenditure school districts. Apparently, according to reliable research, a higher correlation exists between the expenditure level and amount of preparation of the teaching staff than between expenditure level and any other single index of school quality.

Purpose

The primary purpose of this study is to determine the relative differences between the states according to seven criteria, namely: 1) total operating expenses, 2) capital outlay for construction, improvement and teacher's salaries, 3) the cost per public elementary and secondary school pupil, 4) the number of public school elementary and secondary school teachers, 5) the number of public school elementary and secondary school pupils, 6) the teacher-pupil ratio, and finally, 7) the rank of each state using cost per elementary and secondary school pupil as the criterion.

Procedure

The study is employing a reciprocal technique in analyzing the data. For example, the data includes a span of three years, from 1957 to 1960. Twenty-eight states have afforded data for the year 1957-58, eleven states plus the District of Columbia have contributed data for the year 1958-59, and the remaining eight states ' data is for the year of 1959-60. One state, that of Iowa, has been omitted because of incomplete data. It is interesting to note that six of the top ten state's statistics were for the year of 1957-58, whereas three of the low ten states' figures were from the 1958-59 academic year. The source of data for the study was extracted from The Americana Annual 1960, American Corporation, New Yori-Chicago, Washington, D. C. The reader is cautioned at this point as to the limitation of this study, in that is is attempting to compare the states by using three different years, rather than one year, as would be customary in drawing inferences.

Findings and Interpretation

As one may note from Table I, the Middle Atlantic states appear to spend the greatest amount of money for the total operational expenses of their school systems, a mean of $746,383,716. The North Central and Western states rank second and third highest in the nation, total operating expenses suggesting means of $269,462,924 and $183,602,975, respectively. The New England states spend the least, namely, a mean of $86,794,344. In spite of the fact that the New England states spend roughly only one ninth as much as the Middle Atlantic states, a contributing factor, which may in part explain the financial disparity, is that traditionally the New England states emphasize pricate school education to a greater extent than they do public school education. For example, during 1957-58, the Middle Atlantic states were operating some 12,818 public schools, while the New England states were supporting only 2,834 public schools.

Continuing with Table I, we find the South Atlantic states' mean of 31.9 percent expenditure for capital outlay surpassing all other regions. This seems, in part, to verify earlier studies in that the states with the least financial ability are making greater efforts financially to support public schools than do the states with breater financial ability. Following the South Atlantic states, the North Central and Western states rank second and third with a mean capital outlay of 28.9 and 26.2 percent, respectively. The New England area ranks the lowest in the nation, expending a capital outlay of roughly 19.4 percent.

Table I tends to show the greatest concentration of teaching personnel in the Middle Atlantic states with a mean of 76,889.67, whereas the smallest

TABLE I

A REGIONAL COMPARISON OF STATE FINANCES, TEACHERS AND PUPILS IN PUBLIC SCHOOLS*

Regions	Total Operating Expenses	Percent of Capital Outlay[4]	Number of Public Elementary and Secondary School Teachers	Number of Public Elementary and Secondary School Pupils	Cost per Elementary and Secondary School Pupil	State Rank[5]	Regional Teacher-Pupil Ratio
New England:							1 - 24.4
Connecticut[1]	$145, 702, 207	28.1	16, 306	441, 827	$333. 81	22	
Massachusetts[1]	241, 936, 671	18.1	33, 357	781, 343	309. 64	28	
Maine[1]	39, 093, 703	10.1	7, 582	184, 226	231. 24	39	
New Hampshire[1]	29, 498, 944	24.1	3, 959	97, 478	301. 62	29	
Rhode Island[2]	43, 544, 411	8.9	5, 544	130, 510	385. 00	11	
Vermont[1]	20, 990, 129	26.8	3, 117	71, 061	278. 28	34	
Mean	$ 86, 794, 344	19.4	11, 644	284, 497	$306. 60		
Middle Atlantic							1 - 21.4
New Jersey[1]	$363, 644, 280	30.6	42, 073	975, 032	$409. 65	9	
New York[1]	1, 328, 651, 873	22.0	112, 483	2, 092, 162	587. 49	3	
Pennsylvania	546, 854, 995	14.7	76, 113	1, 862, 554	330. 82	23	
Mean	$746, 383, 716	22.4	76, 889	1, 643, 249	$442. 65		
South Atlantic							1 - 28.7
Delaware[1]	$ 27, 842, 085	77.0	3, 213	73, 551	$436. 00	6	
Florida[1]	216, 867, 000	30.1	31. 405	895, 880	291. 78	32	
Georgia[3]	200, 000, 000	30.0	32, 524	981, 223	241. 00	38	
Maryland	160, 790, 542	39.1	20, 561	532, 121	335. 53	21	
N. Carolina[1]	193, 000, 000	25.1	35, 095	1, 060, 187	203. 09	43	
S. Carolina[2]	92, 725, 052	16.7	19, 928	600, 509	177. 10	46	
Virginia[2]	181, 265, 612	28.3	29, 129	822. 250	242. 67	37	
W. Virginia[2]	95, 097, 873	9.3	17, 177	466, 980	222. 95	49	
Mean	$145, 948, 520	31.9	23, 629	679, 087	$268. 77		
South Central							1 - 24.1
Alabama[1]	$114, 998, 229	20.4	25, 881	775, 981	$171. 16	47	
Arkansas[1]	65, 028, 427	5.1	21, 254	420, 011	179. 91	44	
Kentucky[3]	123, 500, 000	34.0	21, 719	614, 607	217. 00	41	
Louisiana[3]	223, 000, 000	30.0	25, 409	701, 000	373. 40	13	
Mississippi[2]	83, 980, 642	48.8	16, 437	569, 832	177. 54	45	
Oklahoma[1]	124, 274, 093	17.1	20, 089	499, 284	270. 09	35	

Tennessee[3]	156,000,000	19.5	26,365	810,500	211.50	42	
Texas[2]	521,375,000	25.8	70,727	1,896,217	299.54	30	
Mean	$176,519,548	25.1	28,485	785,929	$237.52		
North Central							1 - 26.3
Illinois[1]	$580,761,784	31.0	62,989	1,705,867	$410.86	8	
Indiana[2]	265,000,000	31.0	32,659	912,391	312.00	27	
Iowa	(NFA)[6]	(NFA)[6]	(NFA)[6]	(NFA)[6]	(NFA)[6]		
Kansas[1]	127,738,833	28.3	20,884	420,657	774.85	1	
Michigan[1]	502,084,289	35.2	55,254	1,574,606	330.72	24	
Minnesota[1]	223,636,927	53,0	26,538	583,589	364.00	16	
Missouri[1]	304,277,349	(NFA)[6]	16,437	569,832	286.54	33	
Nebraska[1]	72,749,036	22.1	6,484	136,832	299.08	31	
N. Dakota[1]	38,837,419	21.2	6,795	128,950	312.10	26	
Ohio[1]	515,400,609	30.5	66,313	1,793,772	325.08	25	
S. Dakota[3]	46,200,000	8.3	8,097	152,000	345.00	18	
Wisconsin[1]	287.405,924	(NFA)	26,291	653,852	338.39	20	
Mean	$269,462,924	28.9	29,885	784,758	$372.60		
Western							1 - 26.4
Arizona[2]	$ 94,187,337	21.6	21,254	282,466	$389.03	10	
California[1]	1,100,347,465	39.3	104,911	3,027,532	372.91	14	
Colorado[2]	(NFA)[6]	(NFA)[6]	13,939	434,355	(NFA)[6]		
Idaho[1]	35,577,880	42.1	6,096	155,890	259.97	36	
Montana[1]	64,646,030	14.0	6,484	136,832	763.00	2	
Nevada[2]	19,677,541	3.2	2,246	69,092	379.55	12	
N. Mexico[2]	66,077,101	27.0	7,981	201,440	371.59	15	
Oregon[3]	147,500,000	22.3	15,175	383,300	441.00	5	
Utah[3]	74,000,000	35.8	7,720	233,200	345.00	18	
Washington[1]	202,741,402	23.2	19,281	521,159	354.75	17	
Wyoming[3]	31,275,000	33.8	3,574	79,000	450.00	4	
Mean	$183,602,975	26.2	18,969	502,206	$391.77		
District of Columbia	$ 44,330,000	20.1	3,926	109,367	$434.43	7	1 - 27.8

* Source of data: The Americana Annual 1960, American Corporation, New York-Chicago, Washington D. C. Excludes Hawaii and Alaska. Includes the District of Columbia

1 Figures for 1957-58
2 Figures for 1958-59
3 Figures for 1959-60
4 Includes construction, improvement, and teachers' salaries
5 According to cost per public elementary and secondary school pupil
6 NFA means "No figures available."

number of public elementary and secondary school teachers is located in the New England states which show a mean of 11,644.17. The North Central and South Central states rank second and third in this category, with means of 29,885.55 and 28,485.13, respectively.

The largest number of public elementary and secondary school pupils, according to Table I, reside in the Middle Atlantic states, approximately 4,929 748 pupils reported for 1957-58. The New England states' estimate places its number at 1,706,445, the lowest of all regions. Looking at the official enrollment figures from another viewpoint, the Middle Atlantic states show an estimated mean enrollment of 1,643,249.33 pupils per state; the South Central states, a mean of 785,929; the North Central states, a mean of 784,758.91, with the New England states displaying a mean of 284,407.50, or the lowest enrollment figure in the nation.

In comparing the mean cost per public elementary and secondary school pupil, Table I shows that the Middle Atlantic states rank first at $442.65, the Western states second with $391.77 and the North Central states third at $372.60. The South Atlantic states rank the lowest in the nation at $268.77.

The teacher-pupil ratio as illustrated in Table I tends to place the Middle Atlantic states first with a ratio of 1 - 21.4; the South Central states next in line with 1 - 24.4, and the South Atlantic states the greatest teacher-pupil ratio of 1 - 28.7.

Table I illustrates that when the states are compared individually, the combined cost per public elementary and secondary school pupil finds Kansas leading with an expenditure of $774.85, Montana second with $763.00, and New York third with $5 $587.49. At the other extreme, Alabama ranks 47th at $171.16, South Carolina 46th at $177.10, and Mississippi 45th at $177.54. An interesting observation at this point notes that the data for the three highest state expenditures per public elementary and secondary school pupil were for the year of 1957-58, while the three low expenditure states' data were for the year 1958-59 (South Carolina and Mississippi) and 1957-58 for the state of Alabama.

Tables II and III reveal a comparison of the top ten states and the low ten states as to finance and personnel statistics. For example, Table II indicares that the mean cost per public elementary and secondary school in the top ten states amounts to $509.63, while Figure III shows a mean cost of $203.25 for the low ten states. In other words, a top state spends approximately two and a half times more money per public elementary and secondary school pupil than does a low state.

The total operating expense of the top ten states shows a mean of $281,057,722.20, while the low ten states have a mean of $114,468,959.10. As a result, we find a top state operating with a budget two and a half times greater than we find in a low state. The comparison drawn seems to the writer to be a fair and just one, as six of the top states' data are are for the year 1957-58, two states' data are from 1958-59, and only one state's figures are for the year of 1959-60. On the other hand, only two of the low ten states' data are for the year 1957-58, while five states' data are for 1958-59, and the remaining three states' data are for the year of 1959-60. It is the writer's opinion that if the data for the top ten and low ten states could have been compared using the same year for all states, the differences in the total operating expenses would tend to show a wider disparity than is indicated, favoring, of course, the top ten states.

What amount of the budget may be expended for construction, improvement, and teachers' salaries may be determined in part by a state's capital outlay. According to Tables II and III, we may note that the top ten states' mean capital outlay is 27.05 percent as compared to the low states' mean capital outlay of 20.95 percent. A top state's mean capital outlay is approximately six percent more than one would find in a low state. In other words, New York, a top state, during 1957-58, provided a capital outlay of $1,328,651,873, while Maine, a low state, during 1959-60, indicated a $39,093,703 capital outlay. If New York's capital outlay had been for the same year as that of Maine, it is quite possible that the difference would illustrate an even greater margin favoring New York.

The total number of public elementary and secondary school pupils, according to Tables II and III, indicates that Delaware, a top state during 1957-58 had a total of 3,213 teachers, while North Carolina, a low state, had a total of 35,095 during the same year. In combining the top ten states and comparing them with the combined low ten states, we find the top ten showing a mean of 29,205.5 teachers, and the low ten providing a mean of 22,056.7 teachers. Actually then, the typical top state has roughly some 7000 more teachers than a low state.

With regard to the number of public elementary and secondary school pupils, Tables II and III illustrate that New York, a top state, during 1957-58 had the largest total enrollment of pupils, namely, 2,092,162, while North Carolina, a low state, during 1957-58 had a total enrollment figure of 1,060,187 pupils. The mean difference between a top state and a low state may be characterized as negligible as we note a mean enrollment of 625,823.4 pupils for the top ten states as compared to a mean enrollment of 632,508.3 pupils for the low ten states.

The top ten states show a mean teacher-pupil ratio of 1 - 21.4, while the low ten states have a mean teacher-pupil ratio of 1 - 28.7. Apparently, teachers in a top state have fewer pupils per classroom than comparable teachers located in a low state.

TABLE II

FINANCE AND PERSONNEL STATISTICS OF THE TOP TEN STATES

State	Cost per Public Elementary and Secondary School Pupil	Total Operating Expenses	Percent of Capital Outlay	Total Number of Public Elementary and Secondary School Teachers	Total Number of Public Elementary and Secondary School Pupils
1. Kansas[1]	$774.85	$127,738,833	28.3	20,884	420,657
2. Montana[1]	763.00	64,646,030	14.0	6,484	136,832
3. New York[1]	587.49	1,328,651,873	22.0	112,483	2,092,162
4. Wyoming[3]	450.00	31,275,000	33.8	3,574	79.000
5. Oregon[3]	441.00	147,500,000	22.3	15,175	383,300
6. Delaware[1]	436.00	27,842,085	77.0	3,213	73,551
7. Dist. of Columbia	434.43	44,330,000	20.1	3,926	109,367
8. Illinois[1]	410.86	580,761,784	29.1	62,989	1,705,867
9. New Jersey[1]	409.65	363,644,280	30.6	42,073	975,032
10. Arizona[2]	398.03	94,187,337	21.6	21,254	282,466
Total	$5096.31	$2,810,577,222	292.8	292,055	6,258,234
Mean	509.63	281,057,722.20	29.9	29,205.5	625,823.4

1 Figures for 1957-58
2 Figures for 1958-59
3 Figures for 1959-60

4 Total operating expenses derived from local, state and federal funds.
5 Capital outlay for construction improvement and teachers' salaries.

TABLE III

FINANCE AND PERSONNEL STATISTICS OF THE LOWEST TEN STATES

State	Cost per Public Elementary and Secondary School Pupil	Total Operating Expenses	Percent of Capital Outlay	Total Number of Public Elementary and Secondary School Teachers	Total Number of Public Elementary and Secondary School Pupils
1. Georgia[2]	$241.00	$181,265,612	28.3	29,129	822,250
2. Maine[3]	231.24	39,093,703	10.1	7,582	184,226
3. W. Virginia[2]	222.95	95,097,926	9.3	17,177	466,980
4. Kentucky[3]	217.00	123,500,000	34.0	21,719	614,607
5. Tennessee[3]	211.50	156,000,000	19.5	26,365	810,500
6. N. Carolina[1]	203.09	193,000,000	25.1	35,095	1,060,187
7. Arkansas[2]	179.91	65,028,427	5.1	21,254	420,011
8. Mississippi[2]	177.54	83,980,642	48.8	16,437	569,832
9. S. Carolina[2]	177.10	92,725,052	8.9	19,928	600,509
10. Alabama[1]	171.16	114,998,229	20.4	25,881	775,981
Total	$2032.49	$1,144,689,591	209.5	220,567	6,325,083
Mean	203.25	114,468,959.10	20.9	22,056.7	632,508.30

1 Figures for 1957-58
2 Figures for 1958-59
3 Figures for 1959-60

4 Total operating expenses derived from local, state, and federal funds.
5 Capital outlay for construction, improvement, and teachers' salaries.

Conclusions

Relative to the total operating expenses for public schools, it is apparent that the Middle Atlantic states lead the nation, spending an estimated four and a third times more money than the New England states, which rank the lowest-- $2,239,151,148 versus $520,766,065, respectively. The North Central and Western states rank second and third to the Middle Atlantic States.

Capital outlay for construction, improvement and teachers' salaries finds the South Atlantic states spending a mean of approximately $373,044,418.40, thus leading all other regions. The New England states apparently spend the least, roughly a mean of $101,028,616.61.

The North Central states lead the country in the number of public elementary and secondary school teachers, namely, 328,741, while the New England states have the lowest number of similar teachers, some 69,865.

Relative to the total number of public elementary and secondary school pupils, the North Central states surpass all other regions, having an estimated 8,632,348 pupils, while the New England states enroll the lowest number of students, an estimated 1,706,445.

The lowest teacher-pupil ratio in the nation is apparently found in the Middle Atlantic states, namely, 1 - 21.4 (that is, one teacher for every 21.4 pupils). The highest teacher-pupil ratio may be credited to the South Atlantic states, namely 1 - 28.7.

Kansas leads the top three states in the country in cost per public elementary and secondary school pupil, expending $774.85 per pupil, followed by Montana with $763.00 and New York with $587.49. Of the three states which spend the least money per pupil, we note Alabama ranking 47th at $171.16, South Carolina ranking 46th at $177.10, and Mississippi ranking 45th at $177.54.

In comparing the top ten states with the low ten states, we find a top state spending approximately two and a half times more money per pupil as a low state. The top state's mean cost per pupil, estimated at $509.63, compares to the low state's mean cost per pupil of $203.25. In total operating expenses, we note a top state's mean figure of $281,057,722.20, while the low state's mean is 114,468,959.10. The top state's mean capital outlay is nearly nine percent more than a low state, being 29.9 as compared to 20.9 percent. The mean number of teachers per top state is estimated at 29,205.5, while the cimilar figure for a low state is 22,056.7, a difference of 7,148.8 more teachers, favoring the top states. In comparing the total number of pupils in the top states versus the low states, we find the low states leading with 632,508.30 pupils, while the top states' figure is 625,823.40, a difference of 6,684.90, favoring the low state.

If we can assume our earlier premise in the introduction of our study, that pupils attending high-expenditure elementary and secondary schools on the average achieve higher scores in the academic subjects than do pupils attending low-expenditure schools, in addition to the fact that high-expenditure school districts have the money to employ teachers with better preparation than one would find in low-expenditure school districts, we may further infer that the top ten states (Table II) apparently are affording pupils greater opportunities in achieving better scores in academic pursuits than are the low ten states (Table III). Additionally, we might assume that the top ten states, theoretically at least, possess teachers with better preparation than one would likely find in any of the low ten states. Kansas, Montana, New York, Wyoming and Oregon apparently lead the nation in the categories just cited, while Alabama, South Carolina, Mississippi, Arkansas and North Carolina rank the lowest in the country.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

INTERCLASS GROUPING FOR READING INSTRUCTION IN THE MIDDLE GRADES

DONALD ROSS GREEN
Emory University
HAZEL WALKER RILEY
Atlanta Public Schools

THE JOPLIN PLAN (2) is an administrative device for reducing the heterogeneity in reading skill of a group during instruction in reading. A reading period is set aside for pupils in, say, grades 4, 5, and 6, during which these students leave their regular classes and join others in these grades whose reading level is approximately comparable to their own. Thus greater homogeneity in reading skill is attained in each group than is generally found within class groupings. Presumably greater diversity in all other areas also results. Since the regular classroom teacher still provides at least some instruction in reading, most pupils receive instruction from two teachers.

Some years ago Russell (5) reported that the procedure known as "circling" in the San Francisco City Schools did not appear to be particularly effective. Circling is apparently the same as the Joplin plan. Recently a study apparently contradicting Russell's findings has been published by Morgan and Stucker (4).

In the Russell study the subjects in the experimental (circling) group were from one set of schools and the control subjects were from another set of schools considered comparable to the first set. The grade scores in reading of both groups increased about 1.9 in two years (end of grade 4 to end of grade 6).

Morgan and Stucker dealt with the 5th and 6th grades of a single rural school. In each grade a "fast" and a "slow" control class was set up. The experimental and control subjects did not differ in mean IQ, and the pupils were matched on two measures of reading. Retests a year later found the experimental groups scoring substantially higher than the control groups on both tests.

The present study offers data relevant to the questions one might raise because of the differences between the two studies in procedures and outcomes.

Subjects and Procedures

Starting in 1957 a number of schools in Atlanta decided to give the Joplin plan a try. The faculty of each school made the decision independently and implemented it in their own fashion. Four of these schools (hereafter Schools A, B, C, and D) believed to serve a representative cross section of the white population of the city* were selected for study. The experimental procedures followed in each school will be described below.

For various reasons complete data for all the grades considered could not be obtained from all four schools. School A was able to supply data for grades 4 and 6 but not for grade 5; for Schools B and D, 6th-grade data were incomplete but adequate information about grades 4 and 5 were obtained; only the 4th grade of School C could be included.

Control group data were obtained from the records of pupils in these schools and grades for the year prior to the inauguration of the experimental program. Schools A and B first tried the plan in 1958-59. Schools C and D began the preceding year. The experimental group subjects were drawn from those pupils in the appropriate grades in 1958-59. Thus the procedure was new to all teachers and pupils in Schools A and B; it was also new to the 4th-grade pupils but not to the 5th-grade pupils and not to the teachers in School D. Both pupils and teachers in School C had had a year's experience with interclass grouping.

Subjects were matched by school, grade, sex, IQ, and parental occupation. The intelligence test was the Kuhlmann-Anderson, Sixth Edition, Form D, administered at the end of the 3rd grade or the

* The elementary schools are still segregated.

beginning of the 4th grade and, in some cases, again at the end of the 5th grade. Where two scores were available an average was used. The categories used by Warner, Meeker, and Eels (6) were used to rate parental occupation. Because of incomplete records, there were fewer subjects for whom complete data were available among the control groups in each school than among the experimental classes. Each control subject was matched with an experimental subject; then a random sample of 15 male pairs and 15 female pairs was drawn from these pairs in each grade in each school. Mean IQ's of the subjects and median ratings of parental occupation are shown in Table I.

The potential control subjects for whom complete data were not available tended to fall among those with the lower intelligence test scores. Consequently the mean IQ's reported, especially those for School D, are a bit higher than would be representative of the entire school.

The reading scores reported are grade scores (average of two subtests) from the Stanford Achievement Test, 1953 Edition, administered to each group in September or October and again in May using a different form. The Elementary Battery was used with grade 4 subjects and the Intermediate Battery with 5th- and 6th-grade subjects. The mean initial (September) scores for each group are indicated in Table I. No attempt was made to control initial reading scores; the only "significant" differences between the experimental and control groups occurred in the 4th grade at Schools A and C as noted in Table I.

Since each school proceeded independently, the experimental procedure of each will be described.

School A: Each student in the 4th and 5th grades was assigned to one of eight reading groups solely on the basis of his score on the September reading test. The 6th- and 7th-grade students were also divided into eight groups in the same manner. These special reading classes met from 9:15 to 10:00 three times a week throughout the year. The regular classes were kept as heterogeneous as possible within each grade both years. The 16 teachers in the four grades concerned were the same both years.

School B: The students in grades 4, 5, and 6 were divided into ten reading groups. Assignment was based on Stanford Achievement Test reading scores, IQ, teacher judgment, an informal individual inventory and, in some cases, an oral reading test. The testing and group assignments were the responsibility of the school's special reading teacher, who taught an eleventh group of "special problems." In the assignment to groups no child was placed in a group whose level differed from his own actual grade placement by more than two years. Reading classes met daily for 45 minutes. Assignment to regular classes was not based on ability for either the experimental or control groups. The 4th- and 5th-grade experimental students had the same teachers the control pupils had had.

School C: The program was begun with the 3rd and 4th grades in the Fall of 1957 (the year prior to the time the experimental subjects were tested). In 1958-59, 4th-grade enrollment dropped from 65 to 41, so that there was only one instead of two 4th-grade classes. Pupils in grades 3, 4, and 5 were assigned to one of four reading groups on the basis of Stanford and Durrell-Sullivan reading scores and IQ. The fourth and slowest group was taught by the principal. Reading classes met daily for 50 minutes. Regular classes in both years (1956-57 and 1958-59) were entirely heterogeneous within grades. The faculty was reduced in size but not otherwise changed from 1956-57 to 1958-59.

School D: During the year 1958-59 the students in grades 3 through 7 were divided into eight reading groups on the basis of Stanford Achievement Test reading scores. These reading groups met for 45 minutes daily. The plan had been used for three months the previous Spring. Within each grade class assignment has always been made without regard to ability. Unlike the other schools, the teachers in the 4th and 5th grades in 1956-57 had all been replaced by 1958-59.

In all four schools the teacher attempted to use materials appropriate to the reading level of the students in his section. Each school also shifted students from group to group whenever such changes seemed appropriate.

Results

That the experimental groups improved more than the control groups can be seen in Table II. Only the difference between the School D 5th-grade groups failed to be large enough to meet the customary standards of reliability. The gains reported may be compared to an expected gain of roughly 0.7 or 0.8 for groups of average age and intelligence. Thus the control groups gained on the average a bit less than expected, and the experimental groups somewhat more than expected. The differences are quite similar to those reported by Morgan and Stucker.

The differences between gains in each grade of each school have been shown to be large enough to be detected using groups about the size of a typical class. An analysis of variance (not shown) of the 4th-grade gains indicates that the school differences in gains are significant but that the method x schools interaction is negligible. Thus the variation in gain differences is not a matter of concern.

Naturally not all went well with the experimental

TABLE I

MEAN IQ, MEDIAN OCCUPATIONAL STATUS OF PARENTS, AND MEAN INITIAL READING SCORE OF THE SAMPLES

Grade and School	Intelligence Quotient				Parental Occupation				Initial Reading Score			
	Experimental		Control		Experimental		Control		Experimental		Control	
	Mean	S.D.	Mean	S.D.	Md	Range	Md	Range	Mean	S.D.	Mean	S.D.
4A	111.2	7.8	111.4	7.7	2.2	1-4	2.2	1-4	4.62	0.53	5.43[a]	1.20
4B	102.8	10.9	103.4	10.6	4.1	3-6	4.1	3-6	4.25	1.39	4.43	1.28
4C	92.9	12.1	93.1	12.1	5.0	4-6	5.0	4-6	3.56[b]	1.06	3.16	0.78
4D	91.7	10.2	91.8	10.3	6.0	5-7	6.0	5-7	2.96	0.79	3.15	0.81
4 Total	99.7	13.1	100.0	13.1	4.5	1-7	4.5	1-7	3.85	1.18	4.04	1.41
5B	103.5	9.9	103.5	10.8	4.3	3-5	4.3	3-5	5.34	1.22	5.53	1.39
5D	94.4	10.4	95.0	10.6	5.7	3-7	5.6	4-7	3.74	0.75	3.95	1.10
5 Total	99.0	11.1	99.2	11.5	5.0	3-7	5.0	3-7	4.54	1.29	4.74	1.48
6A	111.4	7.5	111.5	7.4	2.0	1-4	2.0	1-4	7.17	1.25	7.19	1.44

[a] $t = 3.19$, $P < .01$ [b] $t = 2.21$, $P < .05$

TABLE II

MEAN READING SCORE GAINS AND GAIN DIFFERENCES

Grade	School	Experimental Group Mean	Experimental Group S.D.	Control Group Mean	Control Group S.D.	Differences Mean	Differences $S_{\bar{D}}$	t	P
4	A	1.55	0.89	0.94	0.78	0.61	0.20	3.01	.01
4	B	1.17	0.66	0.52	0.90	0.65	0.22	2.97	.01
4	C	1.00	0.65	0.49	0.54	0.51	0.16	3.07	.01
4	D	0.96	0.60	0.53	0.68	0.43	0.18	2.38	.05
4 Total		1.17	0.74	0.62	0.76	0.55	0.10	5.75	.001
5	B	1.01	0.68	0.32	0.74	0.69	0.20	3.69	.01
5	D	0.85	0.89	0.49	0.65	0.36	0.20	1.74	.10
5 Total		0.93	0.79	0.41	0.70	0.52	0.14	3.76	.001
6	A	1.28	0.93	0.78	0.88	0.50	0.21	2.38	.05

TABLE III

NUMBERS OF EXPERIMENTAL AND CONTROL SUBJECTS WHOSE GRADE SCORE GAINS FALL INTO THE CATEGORIES UNSATISFACTORY, SATISFACTORY, OR EXCEPTIONAL

Group	Unsatisfactory Gain -1.6 - +0.2	Satisfactory Gain +0.3 - +1.4	Exceptional Gain +1.5 - +4.1	Total
Experimental	24	115	71	210
Control	64	118	28	210
Total	88	233	99	420

$x^2 = 36.9$, df = 3, $P < .001$

groups nor badly with the control groups. If one considers gains up to 0.3 as unsatisfactory progress, and gains of 1.5 or more as exceptional progress, one can see from Table III that about 11% of the experimental subjects made unsatisfactory progress while over 13% of the control group had gains classified as exceptional. A majority of both groups made "satisfactory" progress. Nevertheless the experimental procedures appear to have shifted the entire distribution in a positive direction.

Since initial score was not controlled, the possibility that the gain differences obtained were artifacts of the initial score differences must be considered. It was possible to match half the 4th-grade experimental sample with subjects in the control sample by school, sex, and initial score. The mean initial score of these 60 pairs was 3.66 and their gains were 1.29 and 0.75, yielding a difference of 0.54 ($t = 3.90$, $P < .001$) in favor of the experimental group. This gain difference is practically identical with that of the whole group. Hence it seems unlikely that variations in initial score have distorted the results.

Furthermore intelligence appears to be a more relevant variable than initial score. For the 4th-grade sample the correlations between gain and IQ were +.39 and +.24 for the experimental and control groups respectively; the correlations between gain and initial score were +.18 and -.01 for the experimental and control groups respectively. Hence, where gains are the criterion it would appear more efficient to control intelligence than initial score.

As might be inferred from the results of the analysis of variance, the relationship of gain difference to IQ was considerably less than that for gain and IQ. Mean IQ for each pair and the corresponding gain differences had a correlation of +.13. No relationship between initial score and gain difference could be detected.

It so happened that 15 of the 30 experimental 5th-grade subjects in School B were selected as control subjects for the 4th grade. The gain of these 15 was 0.99 in the 5th grade and 0.47 in the 4th. The difference of 0.52 is less than that found in either grade in this school but is significant ($t = 2.35$, $P < .05$) and corresponds closely to the overall differences in each grade.

Discussion

A great variety of explanations of the above results can be offered. However, since this study paralleled Russell's in design with one exception and obtained results like those of Morgan and Stucker, these explanations should probably not include those suggested by the contrasts between the two previous studies. The selection of control subjects from the same school as the experimental subjects is the one feature of our procedure more like Morgan and Stucker's than Russell's. Therefore, it may be that the two sets of schools used by Russell were not comparable which, if true, could account for the apparent lack of difference he found. The significant differences in gain among the schools studied here points up this possibility.

Of the reasonable explanations remaining only three will be discussed.

1. The most generally accepted rationale for the Joplin plan is the homogeneity in reading skill obtainable. Russell reports that a large majority of the advantages of circling cited by teachers and principals centers around this notion. Morgan and Stucker suggest that this homogeneity permits the slow student to be in a non-threatening situation, thereby increasing the amount of positive reward he can obtain from the material he can actually read. Certainly this is reasonable, and if it is the major factor then the intergrade feature of the plan is merely a concomitant of school size and can be ignored in large schools, as is now done in Schools A and B.

A report by Carlson and Northrup (1) may be taken as supporting this view if one can assume that the pupils in that study were not above average in age, intelligence, or previous training. Also it may be noted that Floyd in his original report of the Joplin plan did not suggest that there was any advantage of the intergrade feature other than the greater homogeneity in reading skill it made possible.

2. The possibility must be considered that a positive element in the situation is the increased heterogeneity, the greater variety of contact and stimulation the multigrade shuffling entails. It is not impossible to conceive of the self-contained classroom as becoming a rather restricted and deadening environment as a school year wears on — same old room, same old teacher, same old bunch of well-known faces, voices, and ideas.

Hull (3) has reported a study from which he argues that the age spread which non-homogeneous, multigrade grouping provides is advantageous to both the younger and the older pupils in such a group. The mean gain differences in reading reported by Hull are considerably smaller than those obtained by Morgan and Stucker and in the present study. Perhaps this means that the Joplin plan combines the advantages of homogeneity and diversity. In any case, a comparison of the intragrade approach with the intergrade procedure would be helpful in evaluating the relative importance of some of these variables.

3. The final possibility to be considered is that because of the special periods more time is spent on direct reading instruction and there is more preparation and emphasis on this instruction by the

teachers than is customary. This special preparation by the teachers, including the procurement of materials, was described both by Floyd and by Morgan and Stucker as a basic feature of the plan. This factor was partially controlled in Morgan and Stucker's study since the teachers of the control classes had access to the additional materials purchased for the experiment. Many of the teachers in the present study cited the necessity of such preparation along with "homogeneity" as an asset of the plan. It should be noted that reading instruction was not (and probably could not be) confined to the special periods. This might mean that the gains were made at the expense of other areas of the curriculum. The comments of several teachers and principals contradict this gloomy notion; in fact, improvement in other areas was claimed. Nevertheless, future studies could profitably include measures of gain in other subjects.

These three explanations, while different, are not mutually exclusive. They all permit the conclusion that some departure from the self-contained classroom has advantages. Therefore, studies of the Joplin plan in departmentalized elementary schools would contribute to the determination of the more influential features of the plan.

Although this study makes it reasonable to assert that most schools would find that their pupils profited from the adoption of the Joplin plan, further confirmation is needed. It would be helpful if future studies would attempt to evaluate the permanence of gain differences as well as the role played by various factors such as those discussed above.

Summary

In four Atlanta elementary schools, the Joplin plan of interclass grouping by reading ability for reading instruction was tried. Typically students in grades 4-5 and 6-7, or 4-6, met daily for 45 minutes in groups made as homogeneous as possible with respect to reading score regardless of class, age, grade, or other abilities. September to May gains in reading scores of 4th-, 5th-, and 6th-grade pairs matched by school, grade, sex, IQ, and parental occupation were compared. Control subjects were pupils in these grades the year immediately preceding the inauguration of the plan. For two schools the procedure was new. The other two schools had used it the preceding year.

The differences in mean gains all favored the experimental groups and ranged from 3.2 months to 6.9 months in grade score. The overall mean difference for the 4th grade was 5.5 months (4 schools, 30 pairs each), 5.1 months for the 5th grade (2 schools, 30 pairs each), and 5.4 months for the 6th grade (1 school, 30 pairs). IQ but not initial score was correlated with gain. Neither variable was related to the difference in gain between the pairs. The advantage of the experimental groups over the control groups did not vary significantly from school to school but the direct gains did.

Brief consideration is given to the way the results may have been affected by three elements characteristic of the Joplin plan: 1)the increased homogeneity of the students in reading ability during reading instruction, 2)the increased heterogeneity of the reading classes in variables other than reading ability, and 3)the greater time, emphasis, and preparation the teachers may devote to reading instruction because of the plan.

Criticism of the self-contained classroom is inferred.

REFERENCES

1. Carlson, E.S., and Northrup, J. "Experiment in Grouping Pupils for Instruction in Reading," National Elementary Principal, XXXV (1955), pp. 53-57.
2. Floyd, C. "Meeting Children's Reading Needs in the Middle Grades: A Preliminary Report," Elementary School Journal, LV (1954), pp. 99-103.
3. Hull, J.H. "Multigrade Teaching," Nation's Schools, LXII (1958), pp. 33-36.
4. Morgan, E.F., and Stucker, G. "The Joplin Plan of Reading vs. a Traditional Method," Journal of Educational Psychology, LI (1960), pp. 69-73.
5. Russell, D.H. "Inter-Class Grouping for Reading Instruction in the Intermediate Grades," Journal of Educational Research, XXXIX (1946), pp. 462-470.
6. Warner, W.L., Meeker, M., and Eells, K. Social Class in America (Chicago: Science Research Associates, 1949).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

RELATIONSHIPS BETWEEN INTELLIGENCE AND REASONING WITH IMMEDIATELY REMEMBERED DISCRETE MATERIALS *

BERT A. GOLDMAN
Mary Washington College of the
University of Virginia

IMMEDIATE MEMORY for discrete symbols and reasoning comprise two processes of the human mind which have challenged psychologists' systematic exploration for years. Scientific investigations have been conducted in order better to understand the existence of these two abilities as separate processes. Positive relationships between intelligence and reasoning with immediately remembered nondiscrete symbols have been found by French and Hunt (1952). However, a controlled investigation of performance on tasks involving both immediate memory for discrete symbols and reasoning with these symbols as immediately remembered has apparently not yet been reported in the literature.

The purpose of the present investigation was to determine the relationships between scores obtained on the Wechsler Adult Intelligence Scale (WAIS) and scores obtained on the Immediate Memory-Reasoning Test (IMRT) designed by the investigator to measure ability to reason with immediately remembered series of discrete symbols. Discrete was used in the sense that there was no apparent relationship within each series of symbols.

It was intended that the results of this study answer two questions. a) Do Ss of superior intellect make significantly higher scores on the IMRT than Ss of average intellect? b) What is the correlation between a group of Ss' IMRT scores and their corresponding WAIS Full Scale scores? In a similar fashion, each of these two questions was asked in terms of WAIS Verbal, WAIS Performance, and the 11 WAIS subtest scores.

Method

Subjects

Forty male, white students between the ages of 16 and 17 years were selected from five high schools within the state of Virginia. Twenty of these Ss composed the group of superior intellect (constituting for the study WAIS Full Scale IQ's greater than 120); the additional 20 Ss comprised the group of average intellect (WAIS Full Scale IQ's between 95 and 105). Initially, 85 Ss were selected on the basis of IQ's reported in school records. Of these, 40 earned WAIS scores which fell within the desired ranges. All students voluntarily participated during non-class hours (i. e., before school, after school, or during study hall).

Materials

WAIS. The 11 subtests of the WAIS were used to determine IQ for each S. Six subtests have been classified as Verbal Scale which includes Information (I), Comprehension (C), Arithmetic (A), Digit Span (D), Similarities (S), and Vocabulary (V). The remaining five subtests have been labeled Performance Scale which consists of Picture Completion (PC), Picture Arrangement (PA), Object Assembly (OA), Block Design (BD), and Digit Symbol (DS).

IMRT. This instrument consisted of four parts each containing 13 items involving either words (Part I), letters (Part II) or geometric figures (Parts III and IV). Each item was composed of two 3 x 5 inch white cards upon which were printed from one to six discrete symbols. Within each of the four parts, each consecutive group of three items contained the same number of symbols. Also, items of each successive group always contained one more symbol than those of the preceeding group. Thus, items 1, 2, and 3 were of equal length while items 4, 5, and 6, also of equal length, contained one more symbol than those of the preceding group. Items 7, 8, 9 and items 10, 11, 12 were similarly constructed. Item 13 contained the maximum number of symbols. Answer sheets were provided for use with test items and with the five sample items which preceded each part of the IMRT.

Intercorrelations among the four parts of the

* Footnote will be found at end of article.

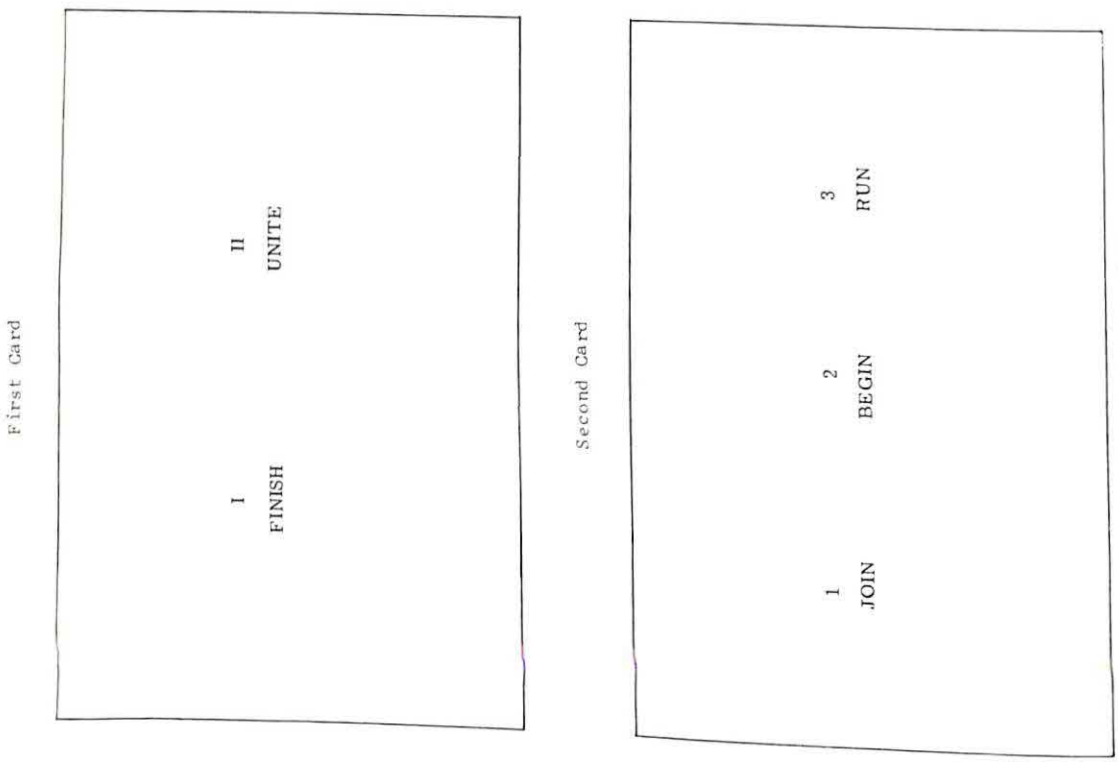


Figure 1. IMRT Part I, Sample Item.

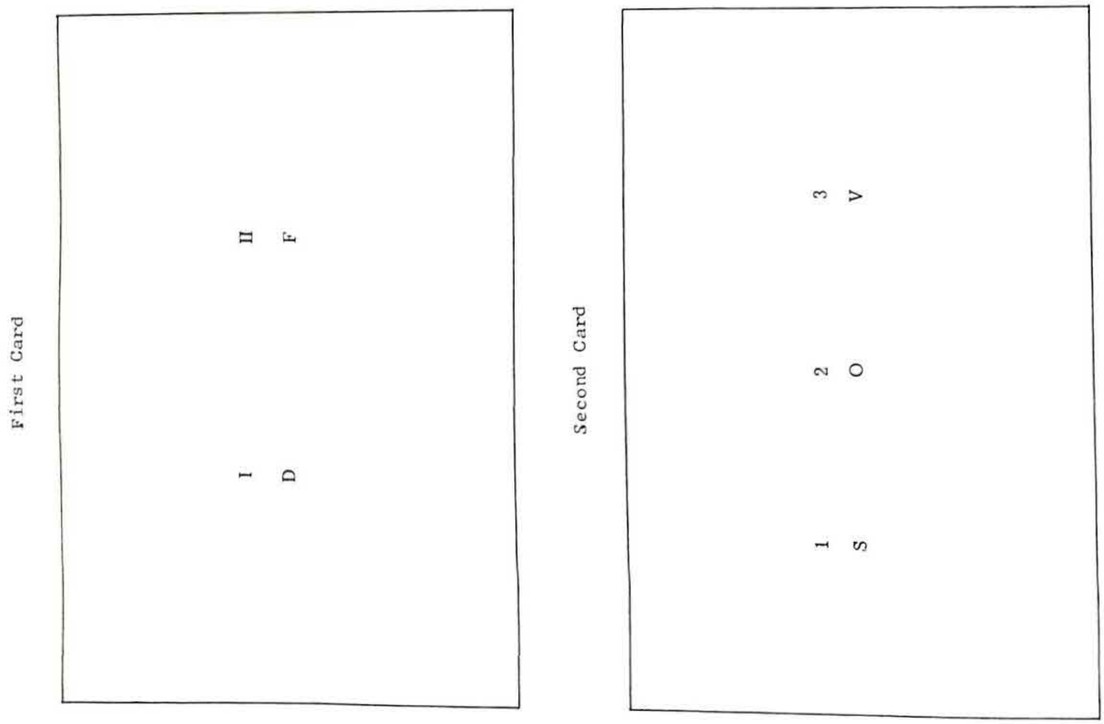


Figure 2. IMRT Part II, Sample Item.

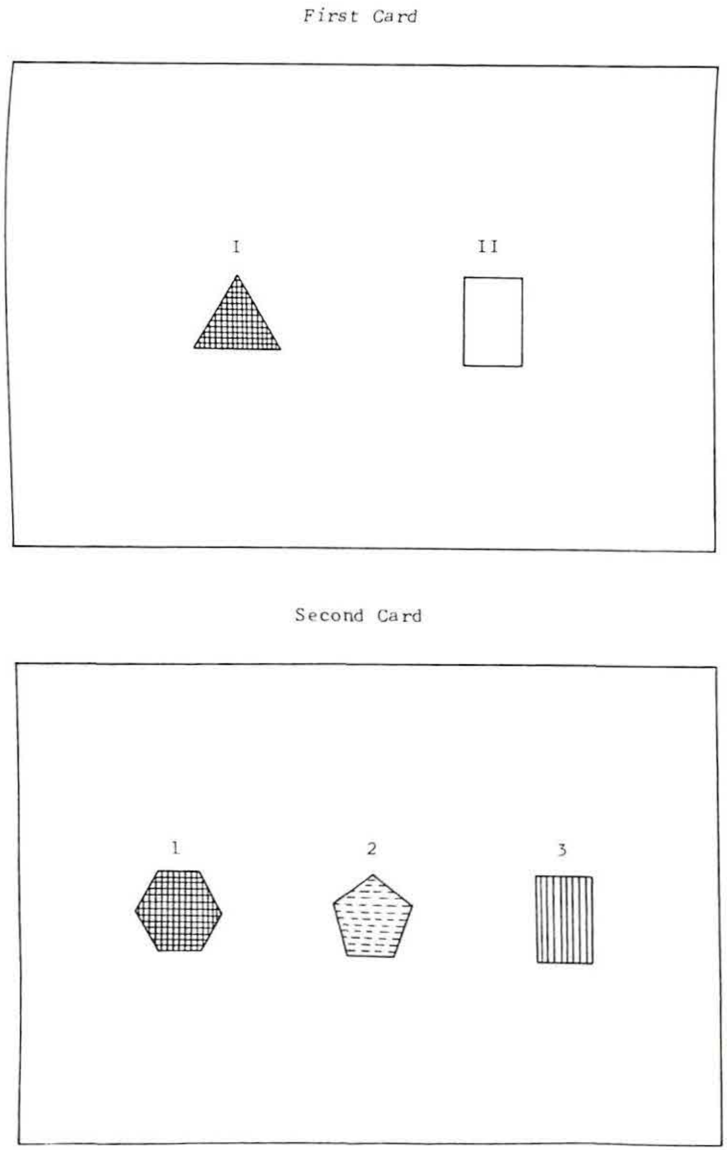


Figure 3. IMRT Part III, Sample Item.

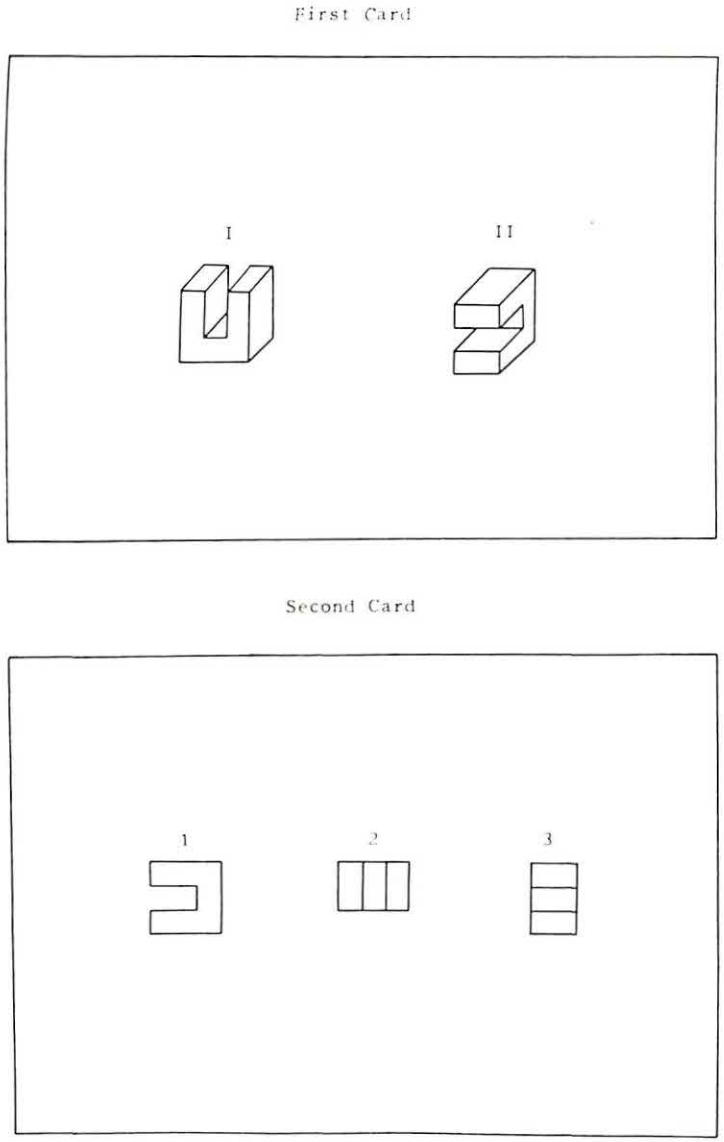


Figure 4. IMRT Part IV, Sample Item.

TABLE I

INTERCORRELATIONS AMONG FOUR PARTS OF IMRT
(Correlations based on 40 cases)

	II	III	IV
I	.59	.62	.59
II		.58	.45
III			.55

TABLE II

CORRELATION OF TOTAL IMRT SCORES WITH CORRESPONDING FULL SCALE AND PARTIAL SCALE WAIS SCORES
(Correlations based on 40 cases)

WAIS	Correlation with Total IMRT
Full Scale	.81
Verbal Scale	.78
Performance Scale	.73
I	.59
C	.72
A	.73
D	.56
S	.64
V	.69
PC	.41*
PA	.43
BD	.41*
OA	.52
DS	.75

Note - All correlations significant beyond .01 level except where indicated.
* Significant at .01 level.

TABLE III

CORRELATION OF PARTIAL IMRT SCORES WITH CORRESPONDING FULL SCALE AND PARTIAL SCALE WAIS SCORES
(Correlations based on 40 cases)

WAIS	IMRT Parts					
	I and II	III and IV	I	II	III	IV
Full Scale	.83	.66	.72	.74	.63	.52
Verbal	.80	.61	.72	.72	.58	.49
Performance	.74	.63	.61	.61	.62	.48

Note - All correlations significant beyond .01 level.

TABLE IV

CORRELATION OF FULL SCALE WAIS SCORES WITH PARTIAL SCALE WAIS SCORES
(Correlations based on 40 cases)

WAIS	Correlation with Full Scale WAIS
Verbal Scale	.96
Performance Scale	.91
I	.85
C	.78
A	.82
D	.74
S	.82
V	.86
PC	.51
PA	.63
BD	.75
OA	.65
DS	.76

Note - All correlations significant beyond .01 level.

IMRT are presented in Table I.

Reliability of the IMRT could not feasibly be dedermined by the split-half technique because of the test's triplicity in design (i. e., in each of the four parts every three consecutive items excluding item 13 were of equal length and contained one more symbol than the preceding triplicate). Therefore, instead of dividing the test items into two groups which is conventional for the split-half method, the items were divided into three groups excluding the four items numbered 13. For example, items 1, 4, 7, etc., from each of the original four parts constituted one section; items 2, 5, 8 etc., from each of the original parts composed a second section; and the remaining items 3, 6, 9, etc., comprised the third section. Correlations among the three parts of this trichotomy were computed (i. e., +.61, +.73, and +.73) which in turn were corrected (because each part represented only one third of the IMRT) by means of the Spearman-Brown formula (i. e., +.82, +.89, and +.89). Finally, from the latter, a mean correlation was determined, using Fisher's Z Function, and was found to be +.87 which was taken as the reliability coefficient.

Viewing Screen. A 29 × 31 inch, black, plywood board equipped with a 4 × 6 inch center, glass window and card frame (which held individual IMRT cards) served as the viewing screen. The purposes of this screen were to provide E with a uniform technique for presenting the IMRT items and to serve as a shield for reducing distractions produced by E in his manipulation of the test materials.

Magnetic Tape and Tape Recorder. A 1/4 inch, 1200 foot magnetic tape (upon which were recorded the directions accompanying each of the four parts of the IMRT) and conventional tape recorder were employed in an effort to reduce variation in presenting the IMRT directions.

Procedure

Approximately one week following the administration of the WAIS, each of the selected 40 Ss received the IMRT. This latter instrument was administered as follows: S was seated before the viewing screen which had been placed in an upright position on a table of writing level height. Enough room on the table between S and the screen was made available for the answer sheet upon which S penciled each answer. Instructions for working the sample problems which preceded each part of the IMRT were played to S from the tape recorder which had been placed in an unobtrusive position. From his seated position behind the screen E displayed both cards of each item consecutively in the screen's glass window. He timed the presentation with a stop watch. S studied the cards from his position in front of the screen. The presentation of each card was preceded by E's anticipatory verbal alert, "Ready!" Viewing time for each card varied from 1 to 18 seconds depending upon the complexity of the item. (e.g., cards containing words or letters were viewed for one second per symbol; those cards containing geometric figures were viewed for three seconds per symbol).

Each symbol on the first card of every item had a Roman numeral printed above it, while each symbol on the second card of every item had an Arabic numeral printed above it. After the second card was removed, S decided which symbol (under Arabic numeral) from the second card was associated with that (under Roman numeral) of the first card. Then S marked on the answer sheet Arabic numerals corresponding to his choice.

The second card of each item contained an extra symbol in order to prevent a forced choice. S was told to respond to all items even though it may have been necessary for him to guess. His score was the total number of correct responses. The entire test required approximately 40 minutes.

A brief description of each of the four parts of the IMRT follows.

Part I consists of a series of words on the first card which is matched from memory according to either sameness or oppositeness in meaning with a series of words on the second card which is remembered also (Figure I). Thus, in Figure I, "BEGIN" is matched with "FINISH" because they are opposite in meaning and "JOIN" is matched with "UNITE" because they have the same meaning.

Part II consists of a series of letters on the first card which is matched from memory according to likeness in sound of either the first or last part of each letter's phonetic reproduction with a series on the second card which is remembered also (Figure 2). Thus, in Figure 2, "V" is matched with "D" because the last part of each letter's phonetic reproduction sounds like "ee" and "S" is matched with "F" because the first the first part of their phonetic sound is "eh".

Part III consists of two dimensional geometric designs on the first card which is matched from memory according to either sameness in shape or in shading with a series of designs on the second card which is remembered also (Figure 3). Thus, in Figure 3 design 1 is matched with design I because they have the same shading and design 3 is matched with design II because they have the same shape.

Part IV consists of a series of three dimensional blocks on the first card which is matched from memory with a series of views of the blocks on the second card which is remembered also. Only top or right side views of each block are matched (Figure 4). Thus, in Figure 4, view 2 is matched with block I because it is the top view of block I and view 1 is matched with block II because it is the right side of block II.

Results

The t test was employed to determine whether the 20 Ss of superior intellect scored significantly

higher on the IMRT than the 20 Ss of average intellect (question a contained under purpose). The t test was also used to determine the significance of difference between IMRT scores which were dichotomized with respect to the 20 highest and the 20 lowest WAIS Verbal Scale, Performance Scale and 11 subtest scores. All t tests with the exception of the one involving data from PA were significant beyond the .01 level of confidence; that from PA was significant at the .05 level. These findings were confirmed by Mann-Whitney U tests, all of which were significant beyond the .001 level of confidence except the one which involved PA. This U test was significant beyond the .01 level.

In order to answer question b of the purpose, coefficients of correlation between the 40 IMRT scores and the corresponding WAIS Full Scale, Verbal Scale, Performance Scale, and 11 subtest scores were determined by the product-moment method and were significant at or greater than the .01 level of confidence (Table II).

Additional correlations were computed between partial IMRT scores and WAIS Full Scale, Verbal Scale, and Performance Scale (Table III). All were significant beyond the .01 level.

For comparison, coefficients were also calculated to determine the correlation between the WAIS Full Scale scores and Verbal Scale, Performance Scale, and 11 subtests. All of these r's were significant beyond the .01 level of confidence (Table IV).

The results of this investigation suggest that ability to reason with immediately remembered discrete symbols inferred to be measured by the IMRT is closely related to the abilities measured by the WAIS which are presumed to be indicative of intelligence. In addition, when Tables II and IV are compared, it is noted that a closer relationship appears to exist between scores from the complete IMRT and those of the WAIS Full Scale (r= +.81) than exists between seven of the 11 WAIS subtests and the WAIS Full Scale. Also, Table III shows that when IMRT scores from parts I and II are combined and correlated with WAIS Full Scale, the resulting coefficient (+.83) is exceeded by only 4 of the 13 coefficients computed between WAIS Full Scale and WAIS partial scores (Table IV).

Although these results are based upon a relatively small, delimited group of subjects the findings are too significant to go unacknowledged. Further investigation is needed to determine the feasibility of using items which combine reasoning with immediate memory as an integral part of tests of general mental ability.

Summary

The primary purpose of this investigation was to determine the relationships between performances on a standardized test of intelligence (i.e., the WAIS) and performances on a battery of items which required the Ss to reason with immediately remembered discrete materials. This latter instrument entitled the Immediate Memory-Reasoning Test (IMRT) was designed by the Investigator for the study and contained 52 items each presented in a viewing screen on separate cards. Data obtained from the IMRT and WAIS scores of 40 eleventh and twelfth grade boys ranging in IQ from 95 to 143 indicated a significantly high degree of relationships (most significant r's included those of +.81 between WAIS Full Scale and Total IMRT and +.83 between WAIS Full Scale and IMRT parts 1 and II combined) between abilities measured by the WAIS and abilities involved in reasoning with immediately remembered discrete materials.

FOOTNOTE

1 This article is based, in part upon the author's doctoral dissertation accepted by the Graduate Division of the School of Education, University of Virginia and conducted under the direction of Virgil S. Ward, Henry E. Garrett, and Richard L. Beard. The author wishes to thank all those who participated in the study.

Portions of this article were presented in a paper read at the American Psychological Association Convention in Chicago, Illinois, September, 1960.

REFERENCE

French, E. G and Hunt, W. A, The Navy North-Western Matrices Test. Journal of Clinical Psychology, 1952, 8, 65-74.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

A STUDY OF ASSOCIATION, REINFORCEMENT, AND TRANSFER IN BEGINNING READING

ANNE PHILLIPS McCREARY
Cornell University

A CHILD, as he enters school, is immediately confronted with the major task of his elementary school career. He must learn to read, not only to read quickly and effectively, absorbing material in a short time, but also to read for meaning and to correlate what he reads with what he has previously read.

The beginning reader must acquire the skills of recognizing printed words, associating a meaningful concept with these printed symbols, and interpreting them in many different sentences. Facile, independent, meaningful reading that brings enjoyment is the teacher's aim.

Although every teacher of reading strives to provide the right type of instruction to develop the necessary skills, independence might profitably be encouraged sooner in first graders. A belief that the three basic processes of association, reinforcement, and transfer could be a key to progress in the development of reading skills and independence prompted the present study. These factors could even be emphasized at the pre-primer level where, in the writer's experience, beginning readers often do not develop or use these processes. Materials and methods designed so that maximum opportunities occur for transfer, reinforcement, and association, if successful, would develop desirable attitudes and substantially ease the teacher load.

From the review of the literature came guides for the design of the materials and their use. The psychology of language and thought (2, 8, 9, 15) and the Gestalt theory of learning (4, 5, 6, 16) offer strong support for the sentence method of presentation. This method of presentation emphasizes meaning and organization which are important in forming associations according to Bartlett (1), Sigel (12), and Staiger (13). The reinforcing effect of knowledge of success and the effect of failure on performance have been pointed out by Sears (11), and Lantz (7).

Reading studies related to association were available but none were reported on transfer or reinforcement in beginning reading. Rivkind (10) found that children differ in their response to different kinds of stimuli. Finally, Vernon's studies (14) and those of Gray (3) suggest the word and the sentence as the unit of presentation for the beginning reader.

The Problem

This study was an attempt to develop experimental materials and to use them in first-grade reading classes during the pre-primer period, in search of evidence to support the belief that:

1. Association, reinforcement, and transfer are important psychological processes in learning to read.
2. Materials based on these principles and deliberately planned to develop such processes can be used profitably as early as the pre-primer stage in reading.
3. Pupils who are skilled in the application of the principles of association, reinforcement, and transfer surmount the problem of learning to read more easily and quickly than do their unskilled counterparts.
4. Use of such materials will not only permit the teacher and the students to increase time-use efficiency but also improve classroom organization.

If such evidence is found, it would enable educators to more effectively evaluate, and make a better choice from the many different methods and materials now in use.

Research Instrument

The instrument has been designed so that opportunities for association, transfer, and reinforcement occur in the following manner:

I. Association
1. visual—The child sees the words and associates them with the picture.
2. visual-auditory—The child sees the words

*This report is part of a dissertation submitted to the Graduate School of Cornell University for the Degree of Doctor of Philosophy, 1961.

and the picture. At the same time he hears the words.

3. visual-verbal—The child sees the words and the picture. At the same time he says the words.
4. meaning—This association builds up through the presentation by the teacher, as the child listens to the story, looks at the picture, and anticipates the climax. It continues as he decides upon the final sentence, says it aloud, and hears other children repeat it. The meanings of some individual words become delimited and more sharply defined as these are singled out or isolated from the context.

II. Reinforcement

1. Whenever association occurs there is a reinforcing affect. Thus, in each example listed under association, reinforcement is also taking place.
2. Repetition, when accompanied by effect, is reinforcing. A child tends to repeat that which is satisfying. Use of this research instrument structures the learning situation.
3. The reinforcing effect of reward is an important factor. Examples of reward occur in the following ways:
   a. Knowledge of Success
      1) During the teaching of the new words the child is assured of success either in anticipating the final sentence or in repeating the sentence after he has heard it, and then in memorizing it successfully.
      2) During the independent reading which may involve either a group situation or individual reading, knowledge of success is present. This is also true in any preparatory or seatwork type of activity. Each time that the child is successful in matching an unknown word with a known word on a picture card and in using this word to get meaning in a new context, reforcement is created within the new context.
   b. Knowledge of Results
      After the initial presentation, this is a self-teaching method, where the children help themselves. They know where to look for assistance and how to discover independently the meaning of the unknown words.
   c. Social Approval
      During the teaching situation, as the child is successful in repeating the new words, he receives social approval from the teacher and from the peer group. As the child uses the picture with other children as a spare time activity he receives approval from the peer group. The ability to read independently brings approval from parents, teacher, and peers.

III. Transfer

1. Transfer takes place as the child matches the word in the new context with the word on the picture card, determines its meaning, transfers this meaning to the new situation, and reorganizes the components of the sentence into a whole.
2. As the child uses this method of self-help in his independent reading, he is transferring the method, using it whenever necessary with different words and different picture cards.

Method of Using the Research Instrument

The teacher presents the picture to the group of children, tells the story, and stops for the last sentence which most of the children will then be able to provide. Picture clues and introductory story context have been designed in such a ways that the desired answer (sentence which is printed under the picture) is invariably given by the child. The children repeat the sentence aloud and actually memorize it. New words are introduced gradually. Nine pictures use only one new word. Nine pictures present two new words. Seven pictures introduce three new words. Twenty-two of the forty-six words appear on more than one picture. Thus the amount of emphasis on words in isolation during the use of the research instrument will vary from picture to picture.

After the actual teaching procedure is completed, the picture card is posted on the bulletin board where the child has complete freedom to refer to it at any time. When a child encounters a word in his independent reading, which looks familiar but which he doesn't remember, he checks the picture cards, finds the word, and by repeating the known sentence, identifies the word which he needs. He then transfers this word to the new situation, reorganizes the components of the new sentence into a whole and makes meaning out of the new group of words. Some children need only glance up from their desks. Others bring their book to the bulletin board and search for the necessary word at close quarters. During free time activities, groups of children often choose to play games which focus on the cards. This usually involves spotting the words or reading the sentences.

Subjects

Three hundred and eight first-grade pupils in one school system were assigned to either a control or an experimental group. They formed three samples which differed in composition. Sample I

included children of all levels of ability from two schools. Sample II was composed of only average and below-average students from several schools. In Sample III, subjects in both the control and experimental groups were from the same school, but there were very few above-average children enrolled. Table I shows the characteristics of these samples.

Procedure

During the pre-primer stage of learning to read, the control group used the method recommended in the basal reading series. The experimental group used the picture-sentence cards in addition. A word recognition test was administered at the end of a two-month period and a standardized reading achievement test three months later. Independence was evaluated in terms of ability to proceed unaided with exercises in the preparatory book. A questionnaire was designed which elicited information about independence, fluency, and enjoyment. In addition, teachers were asked their opinion regarding the helpfulness of the materials, in terms of organizational efficiency, and effective utilization of time. Multiple regression analyses controlled for the following independent variables: sex, intelligence, chronological age, reading readiness, kindergarten attendance, and section within first grade. Analyses were run using results from each of the two tests, for the total group, for each of the three samples, and for the high and low section of one sample. These results are presented in Tables II, III, and IV. Questionnaire data were tabulated and summarized.

Results and Conclusions

Under the conditions of this experiment and on the basis of measures of reading achievement, findings revealed that, on the average, the following was true:

1. For the experimental group and the control group as wholes, children taught by the method prescribed in the basal reading series had higher reading scores than children who used the experimental materials.
2. For experimental and control subjects in the different samples,
    a. in the sample which included children of all ability levels, the subjects who used the experimental materials were greatly superior;
    b. test results for samples in which children of above-average ability were excluded were higher for pupils in the control group. However, when regression analyses were run for the high and low sections separately, non-significant regression coefficients resulted in the high section.
3. Membership in a high section of first grade was correlated positively with reading competence for both control and experimental subjects.
4. Both kindergarten attendance and membership in a low section of first grade were correlated negatively with progress in reading.
5. Variance accounted for by the independent variables was extremely high.

The general trend, then, was for the use of the materials to be more beneficial in classes which included above-average children and least helpful in those with many slow learners.

Under the conditions of this experiment, and on the basis of teacher evaluation obtained from a questionnaire, these statements can be made:

1. Experimental subjects attained independence sooner but did not retain the lead numerically.
2. Children who used the experimental materials enjoyed all reading activities more than those in the control group.
3. Subjects in the control group were rated higher in fluency of reading.
4. In both the experimental and control groups, children in the fast reading groups were consistently superior to those in middle or slow groups in all aspects of reading competency.
5. Teachers, using the material, were of the opinion that it was highly valuable.

The following conclusions can be drawn from these results:

1. Association, reinforcement, and transfer are important psychological processes in learning to read.
2. Materials based on these principles and deliberately planned to develop such processes can be used as early as the pre-primer stage in reading with children who are grouped heterogeneously or with bright children grouped homogeneously.
3. This does facilitate the learning-to-read process.
4. In addition, teacher load is eased as children develop independent work habits earlier and enjoy reading to a greater extent.
5. Whatever material is used, to be labeled as "slow" has detrimental effects on accomplishment.

REFERENCES

1. Bartlett, Sir Fredrick Charles. Remembering (Cambridge: Cambridge University Press, 1932).
2. Buhler, Karl. The Mental Development of the Child (London: Routlege and Kegan

TABLE I

CHARACTERISTICS OF SAMPLES

Characteristic	All Levels: I	All Levels: III*	No Above Average Pupils: II	No Above Average Pupils: IV	No Above Average Pupils: High III	No Above Average Pupils: Low III
Mean Intelligence Quotient	111	99	106	106	100	96
Mean Chronological Age (months)	73	71	72	72	72	71
Mean Reading Readiness Scores	1.2	0.8	1.1	1.0	0.9	0.7

*Within-school study; few above-average students enrolled.

TABLE II

COEFFICIENTS OF REGRESSIONS RELATING WORD RECOGNITION SCORES TO THE INDEPENDENT VARIABLES

Independent Variables	I	II	III	IV	High III	Low III
X_1	-.470 (.771)	-.824 (.795)	2.508 (1.606)	.294 (.634)	1.034 (2.331)	3.968 (2.574)
X_2	6.818 (.834)	-4.278 (.903)	-4.517 (1.551	-1.714 (.639)	-1.694 (2.322)	-7.104 (2.059)
X_3	-.671 (.912)	-1.293 (.988)	-1.183 (2.129)	-.442 (.695)	.023 (5.122)	-1.264 (2.269)
X_4	1.504 (.846)	4.835 (.990)	1.027 (1.899)	2.175 (.686)	....	....
X_5	.159 (.031)	.111 (.040)	.090 (.067)	.137 (.027)	.034 (.101)	.069 (.095)
X_6	.174 (.120)	.212 (.124)	-.018 (.088)	.056 (.064)	-.615 (.304)	-.246 (.365)
X_7	.761 (.733)	4.818 (.935)	8.633 (1.498)	5.876 (.645)	1.09 (.276)	.780 (.174)
R	.81	.77	.75	.68	.73	.84
R^2	.65	.60	.56	.46	.53	.71

Notes: 1. X_1 = 1 if male sex; = 0 if female
X_2 = 1 if in experimental group; = 0 if in control
X_3 = 1 if attended kindergarten; = 0 if not
X_4 = 1 if in high section of grade; = 0 if low section
X_5 = intelligence quotient in units
X_6 = chronological age in months
X_7 = reading readiness grade placement score
2. Numbers in parentheses are the calculated standard errors of the coefficients.
3. R = correlation between actual scores and those predicted by the regression coefficient.
4. R^2 = percentage of variance of the dependent variable accounted for by the independent variables.

TABLE III

COEFFICIENTS OF REGRESSIONS RELATING READING ACHIEVEMENT SCORES TO THE INDEPENDENT VARIABLES

Independent Variables	Sample			
	I	II	III	IV
X_1	.022 (.071)	.081 (.058)	.099 (.082)	.065 (.046)
X_2	.484 (.084)	-.396 (.065)	-.363 (.080)	-.138 (.047)
X_3	-.392 (.091)	-.221 (.071)	-.003 (.109)	-.043 (.051)
X_4	.134 (.085)	-.085 (.072)	.113 (.097)	.039 (.050)
X_5	.041 (.003)	.016 (.003)	.013 (.003)	.018 (.002)
X_6	.049 (.012)	.019 (.009)	.000 (.005)	.009 (.005)
X_7	.089 (.073)	.324 (.068)	.224 (.077)	.319 (.047)
R	.88	.74	.76	.70
R^2	.78	.55	.57	.48

*See notes for Table II.

TABLE IV

TESTS OF SIGNIFICANCE OF REGRESSION COEFFICIENTS

Sample	Dependent Variable	Degrees of Freedom	Regression Coefficient of X_2	Level of Significance
I	(Y_1)	87	6.818	$P < .01$
II		139	-4.278*	$P < .01$
III	Word Recognition	58	-4.517*	$P < .01$
IV	Raw Score	300	-1.714*	$P < .01$
High III		28	-1.694*	N. S.
Low III		24	-7.104*	$P < .01$
I	(Y_2)	87	.484	$P < .01$
II	Reading Achievement	139	- .396*	$P < .01$
III	Grade Placement	58	- .363*	$P < .01$
IV	Score	300	- .138*	$P < .01$

*Control group has the higher value on the dependent variable.

Paul, 1949).
3. Gray, William S. The Teaching of Reading and Writing (New York: UNESCO and Scott, Foresman, 1956).
4. Katona, George. Organizing and Memorizing (New York: Columbia University Press, 1940).
5. Koffka, Kurt. Principles of Gestalt Psychology (New York: Harcourt, Brace, 1935).
6. Köhler, Wolfgang. Gestalt Psychology (New York: H. Liveright, 1929).
7. Lantz, Beatrice. "Some Dynamic Aspects of Success and Failure," Psychological Monographs, No. 59 (1945), pp. 1-40.
8. Maslow, A. H. "Cognition of the Particular and the Generic," Psychological Review, LV (1955), pp. 22-40.
9. Piaget, Jean. The Language and Thought of the Child (New York: Harcourt, Brace, 1926).
10. Rivkind, Harold C. "The Development of a Group Technique in Teaching Word Recognition to Determine Which of Four Methods is Most Effective with Individual Children," Dissertation Abstracts, XIX (1959), pp. 1620-21.
11. Sears, Pauline S. "Levels of Aspiration in Academically Successful and Unsuccessful Children," Journal of Abnormal Social Psychology, XXXV (1940), pp. 498-536.
12. Sigel, Irving E. "The Dominance of Meaning," Journal of Genetic Psychology, LXXXV (1954), pp. 201-7.
13. Staiger, Ralph C. "Meaning Clues to Word Perception," Education, LXXV (1954), pp. 578-81.
14. Vernon, M. D. Backwardness in Reading (Cambridge: Harvard University Press, 1957).
15. Werner, Hans, and Edith Kaplan. The Acquisition of Word Meanings: A Developmental Study, Monograph of the Society for Research in Child Development, Vol. XV, 1950.
16. Wertheimer, Max. Productive Thinking (New York: Harper Brothers, 1959).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

EXPERIMENTAL CURRICULUM FOR THE "MORE ABLE" STUDENTS

RICHARD E. BULLINGTON
Arizona State University

DURING THE MONTH of October, 1957, an experimental program designed to meet the needs of the "more able" students was implemented at Kiva Elementary School, then one of six elementary schools comprising the Scottsdale School District, Scottsdale, Arizona.

A small segment of the parents approached the principal in midsummer of 1957 with the thought that perhaps the neglected group of students at the elementary school level were the bright children. It was their opinion that the typical curriculum met the needs of the large average group with any additional programs or expended effort on the part of the faculty directed primarily to the slow learner. In essence, this small group of patrons voiced the opinion that not much was being done to intellectually challenge and mentally stimulate the "brighter" pupils. This particular group of parents pointed out that the special programs in existence (area-wide) were designed to meet the needs of the slow learner. While these parents realized that the curriculum had to be constructed to meet the educational needs of the large average group and favored special programs created to meet the needs of the slower pupil, they simply raised the question (and challenge): "Should not some effort be made to meet the needs of the brighter children, and if so, how could this be accomplished?" The principal, who was new to the situation at that time, agreed with the fundamental premise of the parents and promised to formulate an experimental program geared to meet the needs of the "more able" students. Following the stimulating meeting with the parents, a study was made to determine what had been done to meet the educational needs of the bright children in the past, and if little had been accomplished, were there plans under way to establish some type of program for the "more able" students?

As this school was relatively new, it was found that no planning had been made in this particular area of consideration. Armed with this information, the principal initiated the development of an "action research" project designed to include the "more able" students attending Kiva Elementary School.

As a result of considerable thinking and study with the faculty, the following broad project purpose or hypothesis was established: To determine whether or not "more able" students would respond to learning experiences different from the standard subject matter offerings and typical teacher presentations promulgated in the regular classroom. In other words, the attempt was made to find out whether or not different learning experiences resulting from the introduction of entirely new subject matter considerations would stimulate and motivate bright boys and girls to expend near maximum mental effort and if such motivation would be characterized by a "deep within" desire on the part of "more able" students to want to learn on the basis of a new and different gain in knowledge.

As a natural and logical outgrowth from such a hypothesis, the following subsidiary purposes were conceived and established as an integral part of the total experimental program:

1. To provide a selected group of "more able students a series of educational experiences designed to be intellectually challenging, mentally stimulating, and intrinsically rewarding.
2. To cover subject matter areas normally not a part of the public school curriculum, yet of such complexity to fully utilize the intellectual powers of the selected experimental group.
3. To devise and construct a curriculum composed of complex and intriguing subject matter

areas so formulated as to allow for significant transfer of previous and present school learnings to the new academic areas of consideration.

4. To determine the ability of this select group to understand, comprehend, and generalize in relation to the types of subject matter areas offered at an advanced level.

5. To subjectively determine the level of motivation exhibited by the students of this experimental group and to discover the effects this unique curriculum had on their performance in the regular classroom.

The paragraphs to follow are descriptive in nature and should give the reader the necessary insight related to the essence and outcomes of the two-year experimental program.

Identification

Students from grades one through eight were selected to participate on the basis of standardized achievement test results, mental ability test scores, teacher observation of academic performance, and pertinent anecdotal notes recorded by the teachers. These factors in a cohesive relation to one another constituted the selection criterion.[1*] In general, those students whose achievement and mental ability test scores fell within the upper ten percent performance levels of their grade-mates (or peers), along with favorable supporting evidence based on teacher observation and anecdotal notes, were selected as participants in the supplementary experimental curriculum. No single item comprising the criterion resulted in a student's selection. An individual was chosen only on the basis of how one single item of the criterion was consistent with each of the other items constituting the criterion. This selected sample represented a total of 105 students. Twelve of this group were eighth graders, ten seventh-grade pupils, fifteen were sixth-grade students, thirteen fifth graders, twelve were selected from the fourth grade, twenty from the third grade, ten were second graders, and thirteen were first-grade pupils.

Preparing the Community

Prior to implementing the experimental curriculum, several meetings were held with parents of all the students who populated the school. The fundamental purpose of these meetings was to explain to the total school patronage the nature of the proposed experimental program, the hypothesis being tested, and the general nature of research studies. These meetings served to prevent misunderstandings on the part of parents and to emphasize the point that only approximately ten percent of the total group of parents would have children involved in the experimental curriculum. These briefings were made primarily to keep the "ninety percent group of parents" whose children would not be involved gain the proper insight as to why only a small segment of the total school population would be selected for participation. It was realized that the initial success of the experimental program would depend upon total patron support. This approach to gain total school-community acceptance of the experimental program proved to be successful.

Program Structure - Upper Grades (5-8)

After proper selection and identification of the students had been accomplished, students were scheduled to meet this special class, simply designated as "X" class, one hour each day. The fifth- and sixth-grade students met together in a combined class during the hour immediately preceding the lunch period. The seventh- and eighth-grade students were combined and assigned to a class which was scheduled during the hour period immediately following lunch.

The regular classroom teachers adjusted their academic schedule so that the subject content missed by the "X" class students was of such nature that it represented no real hardship for these students to make up. Content missed was in the area of spelling, reading, and literature. It should be mentioned that the thinking of the classroom teachers was considered in the formulation of this special program. Consequently, each faculty member fully supported the experimental program and was quite willing to make the necessary schedule modifications to accommodate the experimental program. The principal of the school taught both[2] the morning and afternoon sections of "X" class.

The intelligence quotients for the fifth- and sixth-grade participants ranged from 118 to 145 and in a similar manner the intelligence quotients for the seventh- and eighth-grade pupils selected to participate in the experimental program ranged from 114 to 142. The mean IQ for the fifth- and sixth-grade group was 128.71, while for the seventh- and eighth-grade members of "X" class the mean IQ was found to be 125.32.

In terms of achievement test performance, the spread of average indicated grade-level achievement for the entire group of identified students ranged from a low of 2.8 years to a high of 3.4 years above their (then) current academic grade-placement assignment.

It should also be emphasized at this point that the difficulty level of the various academic areas composing the experimental curriculum was kept high as a means of utilizing the mental powers of

* Footnotes will be found at the end of the article.

the students to the fullest extent and also for the purpose of determining the highest index level of difficulty which these students could still master and comprehend.

Teacher-constructed tests and various evaluation exercises were used to determine the level and degree of student comprehension and understanding. No evaluation techniques were used for the awarding of grades to the students. Grades, as such, had no place in an experimental project of this essential character. Student motivation had to come from the very type of subject matter presented, i. e., interest and motivation based entirely on wanting to gain and "digest" new and intriguing knowledge.

In content areas beyond the scope of the "X" class teacher, aid was solicited in the form of resource specialists from AiResearch Manufacturing Company, General Electric Corporation, Sperry Rand Corporation, and Williams Air Force Base. Several of these corporations furnished pertinent equipment (instructional aids), impossible to secure through normal channels. Advantage was taken of Arizona State Surplus for the acquisition of equipment related to certain aspects of the subject matter covered (e. g., astro compass, telescope, A-12 bubble sextant, plotters, and an assorted array of aircraft instruments). Through the use of resource specialists and the securing of pertinent instructional aids, required learnings were enriched and enhanced.

The Experimental Curriculum

The following areas of unique content were thoroughly covered by the concerned students over a two-year period. The content to be described will be presented in the sequential order taught. It should be remembered that the material presented was covered during a daily period of only one hour and in addition to the regularly prescribed work at each of the given grade levels.

1. Elementary Statistics: This area of study was conceived to be not only unusual but was considered to represent a type of mathematical content which would lend itself to the transfer of previously learned arithmetical skills and concepts. Study in this field meant that a problem-centered approach could be made. With attention given to interpretation of calculated results, the students were induced to think critically. These boys and girls were taught how to construct and interpret the distribution table with emphasis placed on when to use it. The two "X" class groups were taught how to calculate and interpret the mean, median, mode, quartiles, the average deviation, the standard deviation, Pierson's Product Moment Coefficient of Correlation, measures of error, t-test for determination of significance of differences among assorted measures, standardizing scores, and interpretation of the normal curve of probability. Stress was placed on the transfer of previously learned arithmetical processes to this particular academic area plus the developing of proficiency in handling and manipulating all related algebraic formulae.

2. Dead Reckoning Navigation: Using a standard protractor and ruler graduated to one-sixteenth of an inch, the students were taught the fundamental concepts and procedures utilized in dead reckoning navigation. These concepts and procedures included such aspects as magnetic variation, compass deviation, wind drift, true heading, true course, magnetic heading, true airspeed, indicated airspeed, groundspeed, the basic wind triangle, radius of action problems, alternate airport problems, wind correction problems, relative motion, and interception problems. Emphasis was placed on problem-solving situations where the student was simply given pertinent data which he used to geometrically construct the navigational design leading to its solution. Experiences with primary elements of celestial navigation were also provided this special group of students. In addition, the pupils were taught how to use the Weems Plotter and E6-B computer as these tools are related to the solution of air navigational problems.

3. Aircraft Engineering: The students were introduced to the science and mathematics involved in certain aspects of jet and reciprocating engines (e. g., theory of operation and computation of thrust and horsepower), the physiology of high altitude flight, the science of meteorology as related to flying and, as an added attraction, consideration was given to certain elements of aircraft armament and ballistics.

4. Miscellaneous Considerations: The students received a series of lectures in sociology and anthropology with some attention given to the subject of group dynamics. As a dessert, the pupils received a concentrated course in conversational French. The study of conversational French covered a time period of six weeks.

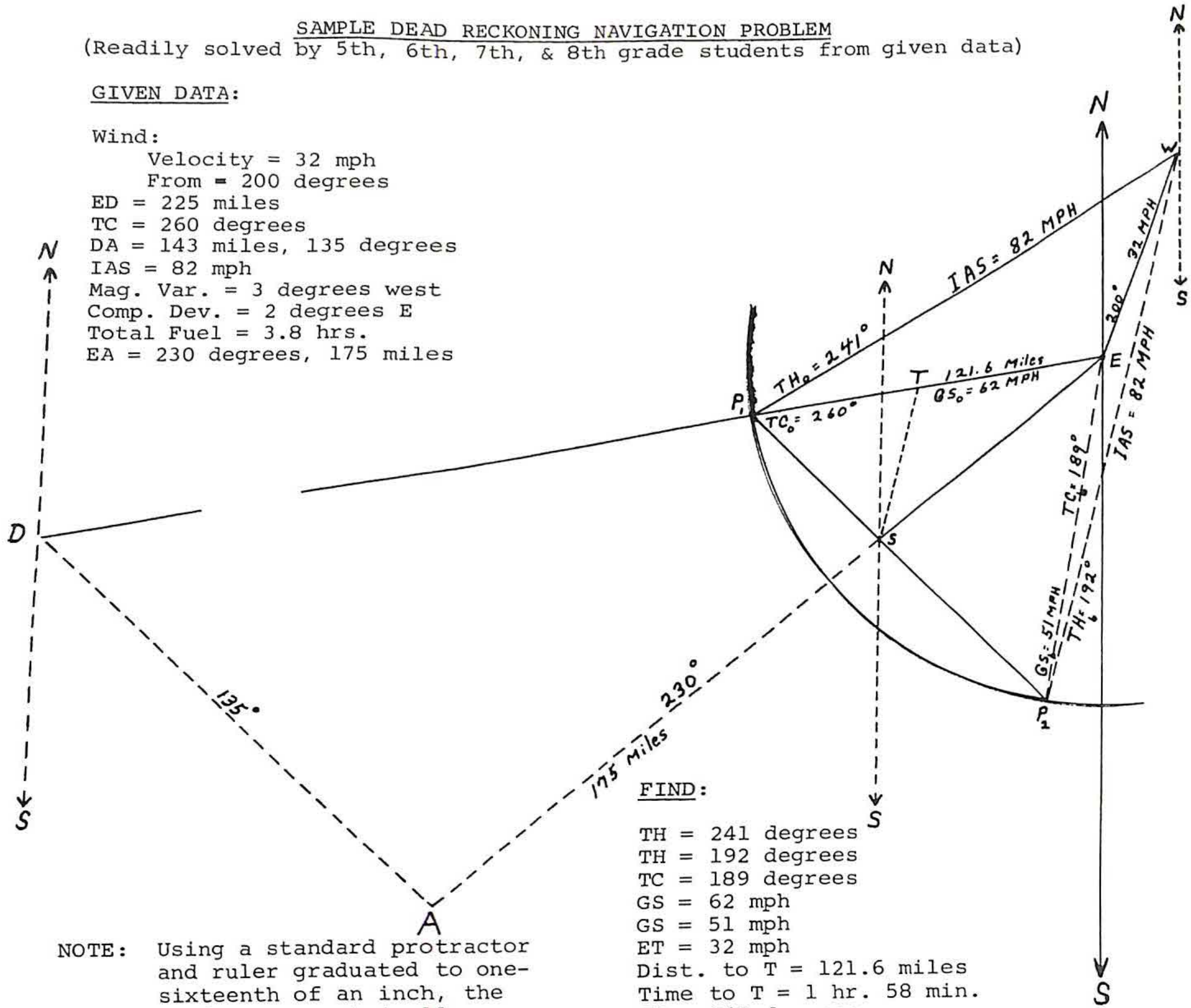
Experimental Curriculum - Lower Grades (1-4)

The students identified as "most able" from grades one through four were divided into two sections designated "Y" groups or classes. The third- and fourth-grade pupils constituted one section of the class and the first- and second-grade boys and girls the second section of "Y" class. The pupils of these two "Y" class sections convened after school for a one-hour period three days each week. These lower-grade pupils were selected in a manner similar to the identification of the upper-grade students.

SAMPLE DEAD RECKONING NAVIGATION PROBLEM
(Readily solved by 5th, 6th, 7th, & 8th grade students from given data)

GIVEN DATA:

Wind:
Velocity = 32 mph
From = 200 degrees
ED = 225 miles
TC = 260 degrees
DA = 143 miles, 135 degrees
IAS = 82 mph
Mag. Var. = 3 degrees west
Comp. Dev. = 2 degrees E
Total Fuel = 3.8 hrs.
EA = 230 degrees, 175 miles



FIND:

TH = 241 degrees
TH = 192 degrees
TC = 189 degrees
GS = 62 mph
GS = 51 mph
ET = 32 mph
Dist. to T = 121.6 miles
Time to T = 1 hr. 58 min.
CH = 242 degrees
CH = 193 degrees

NOTE: Using a standard protractor and ruler graduated to one-sixteenth of an inch, the students geometrically constructed and solved the problem by working only with the GIVEN DATA.

As the lower grades were dismissed at 2:30 p. m., the additional hour did not extend the length of the school day too long for these younger boys and girls. The approach used in providing advanced learning experiences to these pupils was that of modified "team teaching." For the third- and fourth-grade group, two third-grade teachers and one fourth-grade teacher worked together as a group in presenting new and challenging materials to these pupils. The first- and second-grade pupils who made up the second "Y" class section were taught by a team of two first-grade teachers and one second-grade teacher. Upper-grade and district high-school teachers were used as resource persons whenever the occasion warranted the knowledge of subject matter specialists.

Academic Areas Studied

The following specific subject matter content was studied by these two groups of "more able" students; the only significant difference existing between these two "Y" sections was essentially that of content difficulty level. Subject matter areas presented and studied were:

1. Elements of chemistry
2. Fundamental aspects of physics
3. Geology
4. Astronomy
5. Creative writing
6. Literature
7. Special mathematical applications
8. Rocketry

Teacher-constructed tests were administered to measure the degree of student learnings in the considered academic areas. The results of these tests coupled with teacher observation of student performance helped to increase the reliability of teacher evaluation endeavors. No grades were awarded the students who participated in the "Y" classes. Again, the desire to learn had to stem from the very nature of the subject matter presented.

Project Findings

In reference to the initially stated experimental hypothesis and series of ensuing subsidiary project purposes, the data accrued tended strongly to support and substantiate as true this fundamental hypothesis. Furthermore, from close observation and an analysis of evaluation exercises, it was found that the evolving supplementary purposes were each met and consequently fulfilled. The findings are all-inclusive in that they are based on the exhibited performances of the students comprising both sections of the "X" and "Y" classes.

1. Study and analysis of student performance in all the varied subject matter areas covered at the several grade levels, indicate that these select student groups easily grasped and fully comprehended the assorted array of unusual academic areas introduced. The students seemed to accomplish this without undue stress or mental hardship.

2. From close observation of pupil performance and exhibited proficiency in the several content considerations, it seems that the complexity of the "X" and "Y" curriculum was sufficiently difficult to fully challenge and utilize the intellectual powers of the groups.

3. What appeared to be more important than the subject matter itself, was that its uniqueness produced the desirable effect of inducing the participating students to think critically.

4. It was found that the students composing the "X" and "Y" classes readily transferred previous and present learnings to the new and varied subject matter areas studied. This opportunity to transfer previous classroom learnings to interesting and up-to-date problem areas, gave these groups of pupils insight to the "why" for the types of regular and typical classroom learnings and academic requirements. In other words, the students understood the reasons and needs for what had been and was being taught in the regular classroom.

5. Materials presented at advanced levels did not appear in any manner to disturb or frustrate the class members. It was evident that the complexity of the materials covered and the high level of critical thinking required brought into play the near full potential of the student's intellectual power, yet in no way was the level of the subjects studied so high as to be beyond the ability of these students to comprehend and understand. It is felt that as a result of activating the potential mental power of the participating students, these individuals (for this reason alone) thoroughly enjoyed the challenge. Their exhibited performance indicated the existing relationship between their academic accomplishments in the "X" and "Y" classes and the challenge to their intelligence.

6. From observation of the participating students' willingness to expend great effort coupled with amazing accomplishments, it becomes apparent that the motivation of these pupils was at the highest possible level. High motivation on the part of the students induced some parents to complain that they had difficulty in getting their children to go to bed as a result of extreme interest in expanding on areas studied in class. It appears that this exhibited high pupil motivation stems from the type of subject matter offered (i. e., interesting and appealing because it dealt with fascinating things of

the present and the future), the complexity of the subject matter itself, the challenge to their intellectual abilities, and appreciation on their part of the school's recognition of their relative worth and intellectual potential.

7. In most instances, the experiences in the "X" and "Y" classes appeared to stimulate and increase their efforts in the regular classroom. These "X" and "Y" class experiences seemed to act as a "shot-in-the-arm" in relation to their performance in the classroom.

Implications

It is realized that the particular scope of this two-year experimental curriculum is quite restricted and represents only minute segments of possible total subject matter considerations. From experiences gained as a result of working with these students over a two-year period, it would seem that the following implications are worthy of study and reflective thinking on the part of elementary school educators who are interested in meeting the academic and intellectual needs of the "more able" students.

As a result of the insight gained from working with these pupils, it is suggested that special subject matter considerations for bright children at the elementary school level include areas based primarily on the following criteria.

* The subject matter considered should be entirely new and different from what they have had in the past or probably will receive in public school education. The areas covered should be unique, different, difficult, challenging, and interesting in nature. Effort should be made to choose materials which are related to the present and future natural interest areas of the pupils.

* The academic offerings directed toward the "more able" students should be of such character that ready transfer of previous and present classroom learnings can be realized.

* Every effort should be made to prevent the mere duplication of high school subject matter content. There is no need for exact repetition. Duplications of high school subjects at the elementary school level only magnify the problem for the bright students and high-school personnel when the child reaches the secondary level of education.

FOOTNOTES

1. The tests administered were the appropriate forms of the Metropolitan Achievement Test and the Otis Mental Ability Test.
2. The principal's secretary was able to schedule appointments around these two one-hour blocks of time.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

TRAIT IDENTIFICATION AS A MEANS OF PREDICTING ACADEMIC GOAL ATTAINMENT*

REGINA M. GOFF
Morgan State College, Baltimore, Maryland

IT IS THE PURPOSE of this study to consider the feasibility of using personality features as a basis for judgment in predicting academic goal attainment. Interest in the problem has led to intensive investigation (1, 2, 3) of behavior traits which aid in adaptive responses toward academic goals. The attempt hypothesizes the presence of distinctive forces in addition to intellect which are a priori to achievement. Conceptualized, the behavior of emotional energy described as traits is the force which presses toward response. Sthenic,or positive,emotion appears as personal warmth, outgoingness, and the individual is said to possess the trait of sociability. Asthenic, or negative, emotion appears as aggression, belligerence, and the person is said to possess the trait of pugnacity. The assumption concerning the role of behavior features or traits poses inquiry.

A specific question may be stated as follows: Can the relevance or weight of traits in determining academic response be considered in isolation of the complex of mental life and its consequents?

It is assumed that the mind organizes its content, gives meaning to situations and determines which traits shall operate in adaptive response or in carrying out a course of action under conditions of the specified environment. Further, emotional energy, some of which is deeply embedded, has force potential which takes shape under the press of particular environmental stimulations different from the form which it would have taken in the absence of these stimulations. Finally, goal-reaching is the function of variables, dependent upon operation of selected traits which are anticipatory means to ends and which are in turn affected by psychic tendencies and experience.

To the end of examining the question raised and testing hypothesis, the following approach was used. Two measures were obtained on the graduating class and on some members of the freshman class of a liberal arts college. The college selected is a state institution, predominantly Negro, where admittance of graduates of high schools of the state is mandatory though some screening is achieved through use of aptitude and subject-matter tests. Freshmen reported in this study were assumed competent of meeting scholastic requirements of the college. Drop-outs due to finance are relatively few, due to work opportunities and financial aid arrangements. All of the subjects in the study are Negro. The instruments used were the California Personality Inventory and Harrower's Group Rorschach. The senior group consisted of 103 men and 86 women. The original freshman enrollment of this class approximated 600 students. Traditionally at this college, despite original intentions, two-thirds of the freshman group drop out prior to graduation with the smallest percentage of cases attributed to finance. Measures were obtained on the California Personality Inventory for 100 randomly selected freshman men and 100 freshman women. It is assumed that the sample will include potential drop-outs and that trait differentials which relate to motivation may be revealed in comparative examination of freshman and senior responses. Rorschach data were available on a smaller freshman population including 56 men and 25 women.

The Instruments

The California Personality Inventory purports to reveal characteristics of personality "specific to socially functioning individuals" with requisites among other things for academic success. The Inventory consists of 18 scales grouped in four classes or categories. Each category seeks to

*This study was supported by a grant from the Faculty Research Fund of Morgan College, Baltimore, Maryland.

"emphasize some of the psychological and psychometric clusterings which exist among them." Of this number, 10 scales in three of the classes are reported which appear specific to academic intent and behavior. These are:

Class I. Measures of poise, ascendancy and self-assurance. Scales selected: Dominance (Do) - identifies the "aggressive, confident --- persistent, planful individual."

Self-acceptance (Sa) - Assesses factors of independence in thought and action and a sense of personal worth.

Well-being (Wb) - Identifies individuals "relatively free from self-doubt and disillusionment - ambitious and productive."

Class II. Measures of socialization. Scales selected: Responsibility (Re) - identifies the "planful, responsible --- conscientious person --- alert to ethical and moral issues."

Social maturity (So) - Indicates "degree of integrity and rectitude" inclusive of honesty, seriousness, conformity.

Self-control (Sc) - Assesses "adequacy of self-regulation," deliberateness, thoroughness, patience.

Communality (Cm) - Indicates "the degree to which an individual's reactions and responses correspond to the modal (common) pattern established for the inventory."

Class III. Measures of achievement potential. Scales selected: Achievement via conformance (Ac) - "Identifies those factors of interest and motivation which facilitate achievement in any setting where conformance is a positive behavior."

Achievement via independence (Ai) - "Identifies those factors of interest and motivation which facilitate achievement in any setting where autonomy and independence are positive behaviors."

Intellectual efficiency (Ie) - Indicates "personal and intellectual efficiency which the individual has attained."

These scales were selected because components such as foresight, judgment, self-reliance, persistence, industry, effective intelligence and self-direction would appear to be relevant to pursuit of academic goals. Class IV, which includes scales measuring psychological-mindedness, flexibility, and femininity, was not used. It was felt that the aforementioned scales were perhaps most pertinent to the problem.

The study will be presented in the following manner. The question raised will be answered with appropriate data reference followed by conclusions and implications.

Findings

Can the relevance or weight of traits in determining academic responses be considered in isolation of the complex of mental life and its consequents? Trait identification is based on the California Personality Inventory and is shown in Tables I and II. Table I illustrates the mean scores on each of the selected scales for freshman and senior male students. A most striking finding was the great similarity between the two groups. In some instances scores were practically identical. The greatest distance between means was found on the dominance scale with the seniors having a slightly higher mean of 31.3 compared with the freshman score of 28.9. The difference is not significant and the null hypothesis was accepted at the 0.7 level of probability.

Table II represents mean scores of freshman and senior women, where even closer similarity prevailed. The greatest distance between means in this instance was on the achievement via independence (Ai) scale with senior women having a mean of 16.75 and freshman women a score of 15.2 with no significant difference involved. The null hypothesis was accepted at the .05 level of probability.

By way of brief description, the profile for male students was characterized by a slight elevation of two of the three scales in the cluster concerned with poise, ascendancy and self-assurance, and a depression of scores in the area of achievement potential. Highest points were on the dominance and self-acceptance scales indicating uninhibited, aggressive, self-centered behavior with reflections of assurance and confidence. The lower point on the well-being scale, however, the third in this cluster, suggested that this behavior belied the presence of caution, disillusionment, and self-doubt. Conscientious effort and seriousness of purpose indicated on the socialization scale is tempered somewhat by impulsive, changeable, assertive behavior, some lack of control and interest in personal gain as revealed on the responsibility and self-control scales. The communality scale scores, however, indicate that this population does not depart from normal expectations in terms of reliability, adequacy in judgment, and realistic appraisal of situations and would contrast to a restless, imaginative

TABLE I

MEAN SCORES FOR SENIOR AND FRESHMAN MALE STUDENTS ON THE CALIFORNIA PERSONALITY INVENTORY

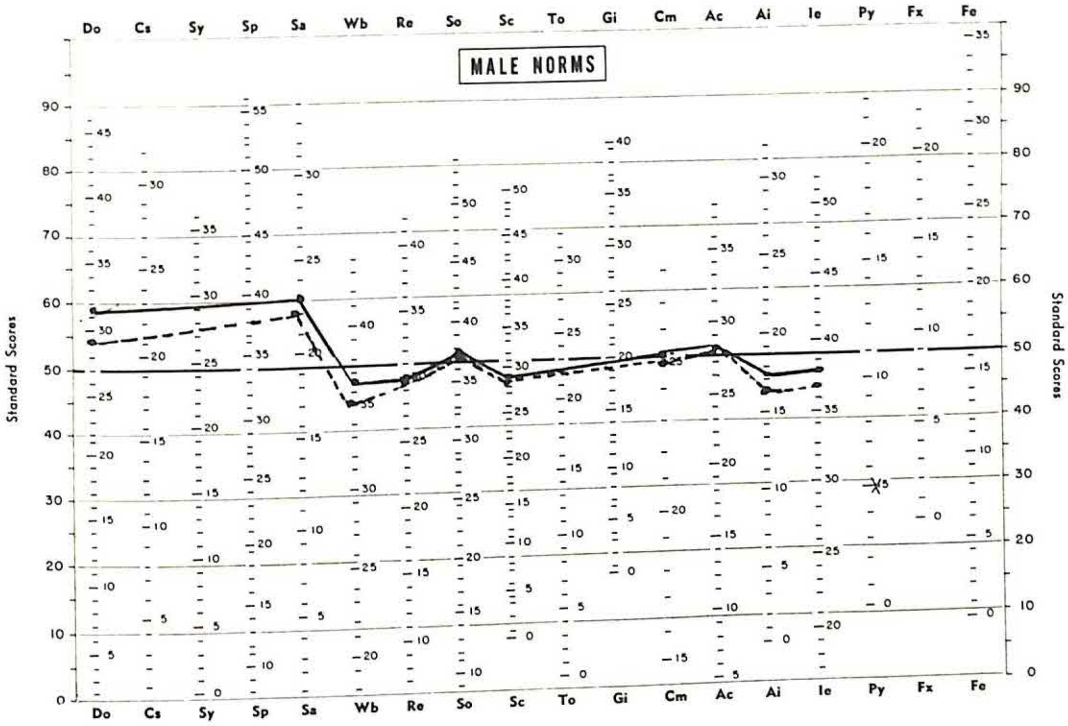
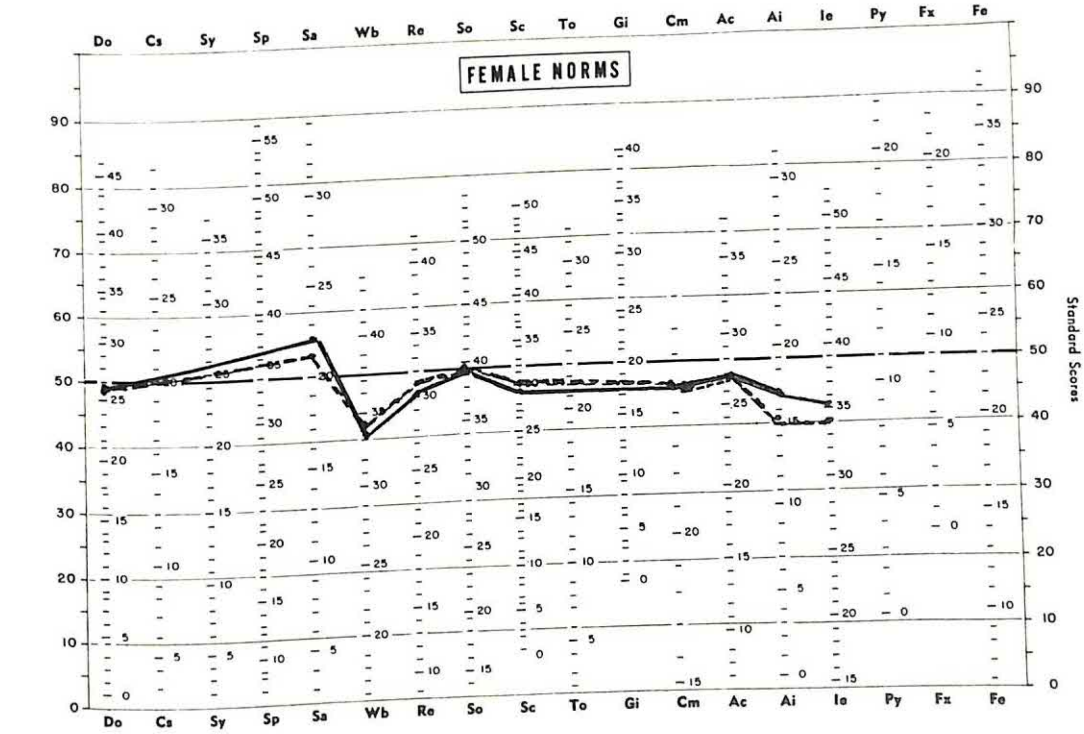


TABLE II

MEAN SCORES FOR SENIOR AND FRESHMAN FEMALE STUDENTS ON THE CALIFORNIA PERSONALITY INVENTORY



Legend for Tables I and II:

Seniors ________________

Freshmen - - - - - - - - - - - - -

group revealing more than expected by way of internal problems and conflicts.

Scores in the achievement potential group were among the lowest in the profile. In terms of these responses, mental capacity is best exercised in traditional, structured settings where creativity, flexibility, and independence in thinking are not demands.

Significantly, the shape of the profile for women approximated that of male students and, if superimposed on the male profile, would show precise directional trends though scores were not as high. Women ranked lower than men on measures of poise, ascendancy, self-assurance, and intellectual efficiency. Men students were either slightly above or at the established norm (found on profile for college students in CPI Manual) on five characteristics out of nine, while women students were below on seven with, in the case of freshman women, two of the seven departing by one standard deviation below the norm.

According to these finds, sex was more related to personality differences than was academic classification.

Trait Differences in Relation to the Intellectual Factor

An attempt was made to see if personality differences would occur between seniors and freshmen if the intelligence factor were held constant. To this end, selections were made of high and low scorers on the intellectual efficiency scale in each of the four groups; freshman and senior men and freshman and senior women. Means were then compared on the remaining traits as illustrated in Table III.

Highest trait scores in all groups with the exception of high-scoring females were on socialization and well-being. Seven groups were thus undifferentiated among themselves and from the student population at large. High-scoring senior women were singularly high on the dominance trait though this, too, is among the cluster which includes well-being. All eight groups were lowest on self-acceptance.

A further comparison of these findings with those of the total population revealed a reversal in the case of two traits. In the total population, the mean on the self-acceptance scale was higher than that on the well-being scale. When high and low scores were isolated, in all instances the Sa score was lower than the Wb score. Intellect may be operative in one's assessment of his personal worth. The data suggest that individuals above or below the average range of mentality are probably more sensitive in comparing the self with others, less confident, and less accepting of the self than are individuals considered intellectually average. The higher score on the Wb scale suggests that feelings of well-being are not affected or lowered in instances where conscientious effort is made toward improvement.

Sex differences were less prominent when the intelligence factor was held constant. Academic classification, again, was not a decisive factor associated with the presence of specific traits.

Rorschach Findings

Rorschach data supported CPI findings. These responses indicate that in the range of psychological experience: intellectual, emotional, inner, and social -- in other words, in terms of cohesiveness -- there was a normal pattern of expression representative of the average of the wider population. Table IV shows that the mean for whole responses in all four groups was 4. The cut-off point for deviant or poor responses was 5.

Table V shows similarities in levels of proficiency among groups when numerical values 1-10 were attached to responses. Scores 1-4 were assigned responses representative of adequate intellectual functioning and inner resources for adjustment; ability to maintain a balance between impulse and values; to delay immediate gratification in deference to long-range goals; to exercise qualities of empathy; and to achieve acceptance of self and harmony in one's social world. Senior and freshman male students had 58% and 60% of the responses, respectively, within this value range while senior and freshman women had 59% and 55%, respectively, in this category.

The low percentage of "1" responses is noticeable, assuming "1" to incorporate all functions of excellence. The range among groups is from 10% to 12% of the total responses. Concurring with intellectual potential scores on the CPI, the factor of cognition, creative organization of thought, is not highly developed.

Predictive Properties of Traits

These findings bring the question raised into sharp focus. Given a set of traits which characterizes an individual or group, can reasonable expectations in response or predictions of academic success be made?

It would appear from the data, revealing great similarity between a group of freshmen (some of whom are potential drop-outs) and seniors who are prospective graduates, that personality features of themselves may not be adequate forecasts of academic goal attainment and that behavior traits might best be considered within the context of the total mental configuration of the individual. The following example is an illustration.

A particular freshman student of the population under consideration holds an original intention of completing requirements for graduation. The student has a great need for status acceptance. The campus environment may or may not be compati-

TABLE III

MEAN SCORES ON TRAITS OF SENIOR AND FRESHMAN MALE AND FEMALE STUDENTS WHEN THE INTELLECTUAL FACTOR WAS HELD CONSTANT

Trait	Male				Female			
	High		Low		High		Low	
	Sr.	Fr.	Sr.	Fr.	Sr.	Fr.	Sr.	Fr.
Do	32	34.2	29.5	27.5	45.1	30	25	24.1
Sa	24.6 (25)	25.3	21.9 (22)	21.1	23.0	22.1	21	20
Wb	38.2	38.4	32.7 (33)	31.8 (32)	38	39	30	31.2
Re	31.6 (32)	31.2	26.7 (27)	27.5	33.5	33.3	29	29.4
So	35.6	34.6 (35)	34.2	32.7 (33)	40	43.8 (44)	34	37.4
Sc	31.0	28.8 (29)	22.7 (23)	24	35.2	38.1	24	26.4

TABLE IV

WHOLE AND WEIGHTED POOR RESPONSES ON RORSCHACH

R	Male		Female	
	Sr.	Fresh.	Sr.	Fresh.
w	4.4	4.4	4.0	4.6
wp	9.12	11.3	9.4	13.5

TABLE V

RORSCHACH RESPONSES WHEN NUMERICAL VALUES ARE ATTACHED

Scores	Male				Female			
	Senior		Freshman		Senior		Freshman	
	No.	%	No.	%	No.	%	No.	%
1	293	12.0	97	12.4	207	10	152	12
2	654	25.4	210	27.0	570	27.5	352	27
3	463	18.0	135	17.2	361	17.4	175	13
4	89	3.0	25	3.1	87	4.2	34	3
5	3		3		1			
6	307	12.0	52	7.0	201	10.2	77	6
7	225	9.0	39	5.0	147	7.1	111	9
8	109	4.0	41	5.1	71	3.4	64	5
9	238	9.2	88	11.2	239	11.0	202	16.16
10	191	7.4	90	11.5	181	8.7	123	9
Total	2572	100	780	99.5	2065	99.5	1290	100

ble to need satisfaction. If it is incompatible, the student may respond to substitute stimuli, in this instance probably a non-college group where aspiration levels are lower and status attainment less difficult. The same set of traits would be operative in the new response for there would exist no variation in motive. However, decreased interest in the academic setting might lead to eventual dropping out. A senior student with a similar trait profile may be observed to have reached stated goals. Incidentally, the illustration also suggests that, apart from personality features, consideration should be given to the place of academic goals in the value hierarchy of the individual, and further, that beyond trait identification, consideration should be given to the relationship of various levels of motivation and to the association process as it operates among energy systems.

The question of trait continuity also arises when predictions are attempted on the basis of trait assessment. A freshman student may identify so closely with an admired upper classman as to lose, at least temporarily, original characteristics. A trait appraisal of such a student at a particular point in time may actually be an evaluation of unrecognized developmental interferences. To some extent, profile similarities between upper and lower classmen shown in this study may be reflections of identification.

It would appear on the basis of the data and in terms of observed behavior that the search for traits which have predictive properties in relation to academic goals is questionable if considered in isolation of this wider context.

Conclusions

Findings on the California Personality Inventory and Harrower's Group Rorschach revealed no significant trait differences between freshmen and graduating seniors. Sex was more related to personality differences than was academic classification. When the intellectual factor was held constant, sex differences were less prominent.

Likenesses among independent groups as indicated by statistical means suggest a universality of traits and the subsequent assumption of modulation of individual uniqueness. In actuality, however, a wide variation in behavior and outcome exists with reference to stated goals. Likenesses may indicate a conscious effort on the part of the individual, at a particular point in time, to submerge personal tendencies in favor of the mass prototype or "student image."

Implications

Test results are very useful in suggesting trends or making generalizations from mass data. However, because of the complexity of ego structure, the probing for distinctive features of individuals which may be used in making sharp predictions is best served through intensive, pluralistic investigation beyond trait identification.

Trait trends should not be ignored. For example, indications noted in this study of self-doubt, self-abasement, and impulsiveness make for vulnerability to substitute stimuli and possible abandonment of the student role. The employment of positive counter-stimuli would minimize, if not offset, their presence and power. The data suggests further that students in general would profit from help in the development of strong, personal convictions which allay vacillation; sound, realistic personal credos; and confidence in decisions bolstered by freedom to act or make adaptive responses in accordance with positively oriented thinking. Guidance of this nature would direct attention to values of merit, increase feelings of worth of self and acceptance of self, and augment the probability of consummation of intent in relation to academic goals.

REFERENCES

1. Froe, O. Non-Intellectual Factors Influencing Academic Achievement, Unpublished Study, Morgan College, Baltimore, Maryland.
2. Gough, H.G. "The Construction of a Personality Scale to Predict Scholastic Achievement," Journal of Applied Psychology, XXXVII (1953), pp. 361-366.
3. Gough, H.G. "Factors Relating to the Academic Achievement of High School Students," Journal of Educational Psychology, XL (1949), pp. 65-78.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

THE PREDICTION OF GAINS IN MEAN PERFORMANCE IN VARIOUS MEASURES OF COMMUNICATION SKILLS RELATIVE TO TYPE OF CURRICULUM PURSUED

ALVIN MARKS
Sonoma State College

ROBERT CATHCART
Los Angeles State College

WILLIAM B. MICHAEL
University of California, Santa Barbara

Background

In the Division of Language Arts at Los Angeles State College, it has been possible for freshman students to follow three curricular plans during an academic year in order to fulfill the requirements in communication experience. Three alternatives have been afforded to students: (1) to take a basic course in English composition during the first semester and then to take a course in speech during the second semester; (2) to reverse the order of courses just mentioned; or (3) to pursue a one-year offering in Language Arts that is intended to represent an integrated or unified approach to communication.

For ease of future reference these three course arrangements may be abbreviated as follows:

ES - English Composition followed by Speech
SE - Speech followed by English Composition
LA - Language Arts— a two-semester course

Purpose

It was the purpose of the investigation to evaluate the relative effectiveness of each of three curricular plans for men and women students at Los Angeles State College during the academic year 1959-60 with respect to four different criterion variables in communication that were intended to represent measures of growth in reading, writing, critical thinking, and listening realized during the school year. Somewhat specifically answers were sought to the following questions:

(1) With respect to each criterion measure of growth was there a significant gain associated with each type of curriculum pursued? In other words, were there reliable average gains in reading, writing, critical thinking, and listening irrespective of the type of curricular plan followed?

(2) With respect to each criterion variable were there significant differences in the amounts of gain in average scores associated with various pairings of courses of study? For example, in the test of reading were there significant differences in the amount of average gain between those three pairs of groups taking the ES and SE, the ES and LA, and the SE and LA offerings?

(3) To what extent was a measure of verbal aptitude related to gains in each of the criterion variables for each type of course of study?

(4) With respect to mean scores in the test of verbal aptitude and with respect to mean gains in the four criterion variables were there significant sex differences?

Subjects and Measures

In Table I, which constitutes a summary of the statistical findings, will be found the numbers of male and female students for whom test data were available relative to each criterion variable and curricular pattern. Unfortunately, not all tests were administered to all classes; nevertheless, it was thought that sufficient data were available for a meaningful study. In the instance of the test in listening, only two curricular arrangements (ES

TABLE I

EIGHT ANALYSES OF COVARIANCE OF GAINS IN SCORES ON FOUR ACHIEVEMENT TESTS CORRESPONGING TO FOUR SETS OF MALE GROUPS AND FOUR SETS OF FEMALE GROUPS WHO PURSUED VARIOUS CURRICULAR PATTERNS ALONG WITH DESCRIPTIVE STATISTICS

Dependent Variable	Curriculum Sequence	N	SCAT VERBAL (Matching Variable) M	SCAT VERBAL σ	Unadjusted Mean Gains in Dependent Variable M	Unadjusted Mean Gains σ	Validity SCAT with Gains in Dependent Variable r	Adjusted Mean Gains in Dependent Variable
Reading (Men)	ES	52	50.50	13.5	3.77	6.5	-.11	3.85
	SE	24	43.96	10.9	4.75	6.8	-.30	4.47
	LA	31	50.42	12.2	1.42	6.5	.03	1.59
Total Group		107	49.01	12.9	3.34	6.9	-.14	(F=3.22)* 2,103 df
Reading (Women)	ES	61	46.28	11.9	4.38	5.9	-.22	4.14
	SE	19	47.21	9.7	7.16	6.4	-.39	7.04
	LA	39	51.33	12.6	3.69	5.8	-.26	4.13
Total Group		119	48.08	12.0	4.60	6.4	-.26**	(F=1.72) 2,115 df
Writing (Men)	ES	71	49.63	11.6	1.72	5.2	-.12	1.70
	SE	24	44.42	10.1	3.33	7.4	.29	3.45
	LA	30	51.20	12.6	1.37	4.9	.25	1.31
Total Group		125	49.01	11.8	1.94	5.7	.02	(F=1.06) 2,121 df
Writing (Women)	ES	95	46.55	11.6	3.78	4.3	-.16	3.74
	SE	19	47.21	9.7	2.74	4.3	-.28	2.73
	LA	38	50.10	13.2	2.26	4.2	.04	2.37
Total Group		152	47.52	11.9	3.27	4.3	-.13	(F=1.55) 2,148 df
Critical Thinking (Men)	ES	38	49.10	13.4	3.58	4.2	-.16	3.61
	SE	28	44.32	11.5	4.96	4.5	.08	4.91
	LA	25	47.52	11.9	5.04	5.5	-.01	5.05
Total Group		91	47.20	12.6	4.41	4.7	-.06	(F= .89) 2,87 df
Critical Thinking (Women)	ES	52	45.37	12.4	4.73	5.6	.16	4.85
	SE	27	51.04	10.8	4.33	4.4	.10	4.19
	LA	31	50.00	13.2	3.97	5.5	.02	3.87
Total Group		110	48.06	12.6	4.42	5.3	.09	(F= .34 2,106 df
Listening (Men)	ES	24	48.17	9.1	2.21	2.9	.24	2.04
	LA	20	53.30	10.2	4.90	6.7	-.33	5.10
Total Group		44	50.50	10.0	3.43	5.2	-.06	(F=3.74) 1,42 df
Listening (Women)	ES	35	47.43	11.4	2.89	4.5	-.38*	2.86
	LA	24	49.17	10.7	4.17	5.4	.30	4.21
Total Group		59	48.14	11.1	3.44	4.8	-.08	(F=1.29) 1,57 df

* Significant at .05 level for the number of degrees of freedom (df) indicated.
** Significant at .01 level.

and LA) could be considered, since the test was not given to any groups pursuing the SE sequence. Moreover, no attempt was made to control the teacher variable or to identify the teachers involved, although approximately 18 teachers were represented.

As to the measures employed, the one serving as a predictor of gains in achievement and as a matching or control variable for equating groups for differences in verbal ability (since groups of differing average linguistic ability could be expected in some instances to show differing average increments in learning), was the Verbal subtest of the School College and Ability Test (SCAT) of the Educational Testing Service. This measure afforded a means for adjusting the amount of gain in the four criterion variables relative to the existence of any differences in the average ability levels of the various groups studied.

The four criterion (predicted or dependent) variables consisted of the standardized tests in Reading, Writing, and Listening in the college-level forms of the Sequential Tests of Educational Progress (STEP) and a standardized test Critical Thinking from the American Council on Education that is intended to measure the ability to exercise critical judgment and to make analytical discriminations. All five tests which had been standardized by their test publishers probably represent about as reliable measures for the functions measured as can be found among various competing tests that are commercially available. In most instances reliability-coefficients (including characteristics of consistency and stability in the scale), equal to or slightly in excess of .90 could be expected.

Experimental Procedure

In addition to a routine administration of the SCAT Verbal test in September for all incoming freshmen equivalent forms of each of the four tests constituting the criterion variables were administered during September 1959 and again during June 1960 in most of the freshmen classes that were pursuing one of the three curricular patterns described. Whenever one of the tests was not given twice the cases were lost for statistical treatment. In those instances in which different (but functionally equivalent) forms of the same test were administered in September and in June, slight adjustments in raw scores were effected from the normative tables of the test manuals in order to ensure comparability of results.

Statistical Procedures

The basic statistical model employed was that of analysis of covariance as described in Dixon and Massey (1957) and in Walker and Lev (1953). Although descriptive statistics (means and standard deviations) were found to indicate average standing and variability in scores on the SCAT Verbal test as well as the average amount of and variability in gains on each of the criterion measures, the analysis-of-covariance procedure was followed in order to obtain net or 'adjusted' mean gains in performance in the measures of reading, writing, critical thinking, and listening that were not associated with differences in the initial level of verbal ability for each of the samples of a given sex following a specified curricular sequence.

Within each sex group the existence of possible significant differences in adjusted mean gains on each test measure relative to the three (or two) curricular patterns was evaluated by an appropriate significance test (the F-ratio) which is indicated in the last column of Table I. Additional significance tests were made for differences between the means of sexes on each of the same achievement tests as well as on the aptitude measure relative to the pursuit of the same curricular pattern.

Findings

The principal findings in Table I may be summarized as follows:

(1) Irrespective of whatever its initial average level of verbal aptitude was, every one of the groups registered a statistically significant (unadjusted) gain in mean level of performance on each of the criterion variables for each type of curriculum pursued. However, in the absence of a control group not receiving any formal instruction in language arts, speech, or English composition, it cannot be affirmatively inferred that the amounts of gain which ranged between 1.42 and 7.16 points on various tests, are necessarily associated with formal instructional procedures, since a portion of each gain might possibly be attributed to incidental learning, increased maturity, or practice effects induced from the initial September experience with the tests. Nevertheless, it is encouraging to note that each group without exception showed a significant gain in average performances

(2) With the exception of the group of men who took the test in Reading there were no significant differences in the amount of adjusted, or net, mean gain in performance relative to the type of curriculum pursued. In other words, there was no definitive evidence to indicate that any one curricular plan was associated with greater average gains in writing, critical thinking, or listening than was any other. Nor was there any systematic trend as to the superiority of one type of curriculum relative to any other. Only in the instance of men upon the test of Reading was there a suggestion of a significant differential in mean gains in terms of a

particular curricular plan pursued. (It appeared that the Language Arts curricular offering was associated with a slightly smaller increment in mean gain than were the two sequences involving Speech and English Composition.)

(3) As a predictor of gains in achievement, the scores of the SCAT Verbal test were for the most part lacking in significant validities. The slight trend for the appearance of negative validities suggests that students of inferior verbal ability may have profited slightly more from the instruction than those students of superior verbal ability. Of course, it should be realized that the attenuation of reliability in difference scores (representing gains) would in part contribute to the finding of relatively low validities for the SCAT Verbal test. In other words from a psychometric standpoint much higher degrees of relationship could have been expected between scores on the aptitude variable and measures of either initial standing in September or final standing in June on each of the four criterion variables than those degrees of validity found between these same scores on the aptitude test and difference scores on the criterion variables.

(4) With respect to neither mean scores in verbal aptitude nor (unadjusted) mean scores in the measures of reading, writing, critical thinking, and listening were there any significant sex differences when each sex group was exposed to the same curricular pattern. Thus, with the presence of comparable ability for two sexes, there were similar gains in the various achievement measures when each sex group took the same type of curriculum. (None of the statistical tests for significance of sex differences yielded probability values of an observed difference in means occuring by chance as low as .05.)

Summary

For several groups of men and of women students who pursued during the 1959-60 academic year at Los Angeles State College, three types of curricula in language arts — a traditional course in English composition during the Fall semester followed by one in speech during the Spring semester, a course in speech during the Fall followed by one in English composition during Spring, and an integrated year-course in Language Arts — measures on equivalent forms of standardized achievement tests in reading, writing, critical thinking, and listening were obtained during September 1959 and June 1960. Statistically reliable gains in mean scores on all tests were attained for all groups of students (each group being members of one sex from all classes pursuing a given curriculum), although the increments could have been attributed at least in part to incidental learning, to practice effects in testing, and to increased maturity as well as to the formal instructional experience. When gains in mean scores on the achievement tests were adjusted for differences in the ability levels of the several groups from knowledge furnished by scores on a verbal aptitude test, it did not appear that any one of the three particular types of curricular practices was superior to any other in terms of mean gains realized on the four achievement tests. In the main, statistically significant validities were lacking between verbal aptitude scores and gains in measures of achievement irrespective of the type of curricular plan pursued. There were no significant sex differences with respect to either average verbal aptitude score or mean score gains in each of the four achievement measures when each sex group pursued the same type of curricular offering.

REFERENCES

1. Dixon, W. A., and Massey, Frank J., Jr. Introduction to Statistical Analysis, second edition (New York: McGraw-Hill Book Company, 1957).
2. Walker, H. M., and Lev, Joseph. Statistical Inference (New York: Henry Holt and Company, 1953).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

THE PREDICTION OF TEACHING COMPETENCE

MAURICE F. FREEHILL
Bellingham, Washington

THE EARLY APPRAISAL of teaching potential is endorsed by both humane and professional considerations but making these assessments continues to be a very difficult assignment. Gross inadequacies in intelligence, subject matter or personality may eliminate some teacher candidates while less obvious limitations, singly or in combination, result in defective or marginally defective teaching. These limitations are particularly important if evaluations and recommendations are to distinguish between degrees of competence and not just between competence and incompetence.

The evaluation of teaching is difficult because teaching involves an unusually wide range of personal qualities and skills and because teaching outcomes must be observed in delayed and inferential fashion. Criteria of success such as income level, length of occupational survival or publications record (used in studies of some professional, managerial and scholarly occupations), are not particularly appropriate for teachers. In practical terms, the criterion of success is the evaluation made by employing schools.

The shortcomings of supervisory reports are many. They are made on brief observation, on tangentially related behavior and on a scale without an established mid-point or range. They may overemphasize classroom control, pleasing appearance or parental approval. The standards are unclear and the teachers are appraised in teaching positions unlike those for which they were recommended. Despite all this, it is practically important to find the range of agreement between college recommendations and field evaluations and secondly, to find which of the college records may be most effectively used to predict teaching success.

These questions were investigated at Western Washington College of Education. A faculty member (specialist in evaluation and unacquainted with any of the teachers or any of the raters), searched the appointment office files to prepare a list of 60 teacher-graduates from the previous year, ranging from "Western's Best" to "Western's Poorest." Such a selection process seemed appropriate since it approximates the way an employer searches for new teachers.

The Relationship Between Recommendations and College Records

Following selection of 60 subjects, "routinely recorded" data were examined to find any which might predict the nature of appointment office recommendations.

First examined were entrance test data. These included an L (Linguistic) and Q (Quantitative) score from an academic aptitude examination, three English scores, reading, mechanics and effectiveness and 10 scores from instruments developed in the American Council on Education Study of Evaluation. These latter included scores on critical thinking, critical thinking in specified subjects and attitude scores. All 14 relationships proved positive but most were below acceptable levels of significance. (This should be expected since the range of scores was attenuated because the ratings cover only students who are graduated and recommended.) The L (Linguistic Aptitude) neared significance and, in general, faculty recommendations had a low positive relation with competence as measured at college entrance but only "Democratic" attitude scores from the Problems in Human Relations Test were truly predictive.

The second group of data came from the completed college academic record. Grade point averages were computed for eight subject matter groupings. When related to the college recommendations, six of these categories (Art-Music, Industrial Arts, Home Economics, Psychology-Education, Mathematics and English-Speech), were significant above one percent probability of error. Two other categories fell below that level with Physical Education between the two and five percent level and Library just above the five percent probability. It may be assumed that faculty recommendations were consistently and closely tied to student success in formal college courses.

A third record was reviewed. It was a report on the social and community life of the student obtained from the offices of Dean of Men and Dean of

Women. The 30 "strong" candidates had been much more active in college affairs. On a 10-point scale they ranked 7.24 compared with 5.74 for the 30 "weaker" candidates. No member of the "strong" group was ever brought before a college disciplinary committee. Likewise, no one in this group was rated as personally unpleasant although two of the "weaker" group were so rated.

Naturally, the ranking of subjects was largely on ratings made by faculty and especially supervisors of student teaching. A second independent rater read each student-teaching recommendation separating the comments into two categories, "professional qualities" and "personal qualities." Academic faculty recommendations were sorted into two classes, those based on class contacts and those based on extra-curricular contacts. Ratings made by faculty who had been on-campus employers of part-time student workers and ratings made by the appointments officer did not permit separation into sub-categories. Thus, there were seven groups of faculty ratings plus the composite rating of the placement officer. As expected, each series of ratings conforms closely with the over-all estimate made by the investigator who selected the cases, falling below the one percent level of error probability in only one case. This was for a group of ratings on "personal qualities" as rated by supervisors of the first student teaching. It seems probable that personality defects are likely to be manifested or heightened when a student is first put in a responsible relation with children. Student teaching supervisors may be particularly observant of these characteristics or their observations may come from a crisis period. They may indeed discover qualities which are central to long-run success or simply defects reduced by subsequent experiences and developing maturity. In any case, these recommendations are most unique of all those on file.

Student personnel files were reviewed for other data with predictive potential. The search was comparatively futile. The counselors' comments were dichotomized as "approving" or "neutral or negative." Counselor approval was significantly related to the "appointment-office" evaluation but not significantly related to evaluations by other faculty nor to ratings from the field. Students who had been employed on campus were significantly more highly recommended by both student teaching supervisors and academic faculty but this did not reflect in "placement-office" appraisal nor was it related to later evaluation in the schools. No significant differences were found between those students who entered college with a clear-cut vocational plan and those who had no plan. Likewise, no significant difference was found between those who entered to become secondary teachers and those who originally chose elementary. The profile from the Kuder Interest Test was examined statistically on the assumption that there might be a profile typical of the successful teacher, but none was found.

In general, the study showed that an employer could search appointment office files and reliably rank candidates in the order assigned by faculty. The superior ratings would go to graduates who had been socially active, personally approved, part-time employed and academically successful in college. There is no evidence that these students came to college with a well defined pattern of interests or a well established vocational plan. The more highly recommended group did slightly better on entrance tests but were clearly marked at college entrance only by a significantly high score on "Democratic" attitude.

Early Field Evaluation

Principals, superintendents and supervisors or vice-principals were asked to evaluate the graduates in February of the first teaching year. Separate ratings were made on "professional" and "personal" qualities. This made six categories of rating. These agreed above one percent probability of error although principals and supervisors were in closer agreement than either principals and superintendents or supervisors and superintendents. The principals and supervisors also more nearly agreed with the college rating which may only indicate that these two groups knew the teachers more than did the superintendents.

Table I summarizes correlations between ratings. The coefficients assume increase significance because each scale is abbreviated by the nature of the group (confined to those able to complete college, interested in teaching, successful in securing a position, etc.). Part 1 of Table I shows comparatively high agreement between various raters on "personal" qualities. Notably, the ratings from a second period of student teaching are lower in agreement with employer evaluation than ratings from an earlier student teaching. This occasions some surprise since the second student teaching is longer, more intensive and temporally closer to the employment period. It may be that the second student teaching evaluation on "personal" qualities discriminates beyond the level which functions in employer evaluations.

Superintendents' evaluations are "reflected" and the agreement with principals should be expected. Beyond this reflection, the superintendents' ratings tend to agree with on-campus employers and be little related to early measures of ability.

The correlations in Part 2 on "professional" qualities are slightly lower than those on "personal" qualities. Once again the rating from second student teaching is comparatively low with the principals' rating. This fact leads to the question of whether or not student teaching supervisors encour-

TABLE I

Part 1 - "Personal" Qualities

	1	2	3	4	5	6	7	8	9	10	11	12	13
1.	1.00	.75	.67	.61	.55	.45	.54	.68	.69	.14	.28	.08	.41
2.		1.00	.79	.53	.29	.45	.53	.60	.66	.13	.26	.02	.18
3.			1.00	.59	.29	.39	.37	.38	.73	.04	.08	-.14	.21
4.				1.00	.73	.71	.35	.64	.64	.23	.42	.16	.46
5.					1.00	.50	.37	.55	.48	.18	.34	.06	.35

Part 2 - "Professional" Qualities

	1	2	3	4	5	6	7	8	9	10	11	12	13
1.	1.00	.72	.66	.52	.70	.45	.54	.68	.69	.14	.28	.08	.41
2.		1.00	.70	.45	.30	.44	.45	.49	.47	-.20	.29	.08	.15
3.			1.00	.44	.42	.50	.35	.50	.51	-.05	.14	.03	.27
4.				1.00	.73	.50	.29	.49	.31	.27	.45	.29	.43
5.					1.00	.54	.38	.68	.36	.25	.40	.25	.48

Part 3 - Over-all Rating

	6	7	8	9	10	11	12	13
6.	1.00	.35	.53	.59	.20	.34	.11	.39
7.		1.00	.59	.61	.14	.33	.17	.48
8.			1.00	.65	.16	.35	.12	.44
9.				1.00	.03	.03	.21	.06
10.					1.00	.50	.46	.06
11.						1.00	.79	.30
12.							1.00	.33
13.								1.00

Key
1. Placement Office rating
2. Principal's rating
3. Superintendent's rating
4. Student Tchg. I rating
5. Student Tchg. II rating
6. Fac. Academic 1 rating
7. Fac. Academic 2 rating
8. Extra-Curr. rating
9. Student Employer rating
10. A.C.E., Q. score
11. A.C.E., L. score
12. N-D, Rdg. score
13. High School G.P.A.

TABLE II

Part 1

	1	2	3	4	5	6	7	8	9
1.	1.00	-.00	.26	.43	.33	.10	.28	.13	.16
2.		1.00	.30	.19	-.12	.25	.31	.36	.01
3.			1.00	.39	.67	.65	.71	.85	.58
4.				1.00	.24	.15	.57	.28	.50
5.					1.00	.41	.43	.50	.45
6.						1.00	.61	.72	.47
7.							1.00	.60	.77
8.								1.00	.72
9.									1.00

Part 2

	1	2	3	4	5	6	7	8	9
Personal	.23	.63	.69	.42	.32	.65	.62	.64	.46
Professional	.56	.26	.66	.63	.37	.55	.80	.74	.81

Key
1. Appearance
2. Health
3. Industry
4. General Culture
5. Social Attitudes
6. Leadership
7. Intelligence
8. Professional Knowledge
9. Valuable Experience

TABLE III

	"Professional"		"Personal"	
	1st year	5th year	1st year	5th year
Principal (first year)	1.00	.43	1.00	.37
Placement Office	.72	.64	.75	.41
Superintendent	.70	.36	.67	.43
Student Teaching I	.45	.59	.61	.46
Student Teaching II	.30	.59	.55	.29
Fac. Academic I	.44	.55	.45	.37
Fac. Academic II	.45	.52	.54	-.18
Fac. Extra-Curricular	.49	.50	.68	.07
Student Employer	.47	.03	.69	-.02
ACE Q Score	-.20	.19	.14	.24
ACE L Score	.29	.42	.28	.26
N-D Rdg. Score	.08	.31	.08	.03

age and evaluate behaviors or qualities which are not the same as those valued and observed either in the college classroom or in the field teaching. On the other hand, the greater agreement between academic faculty and field raters may reflect superficiality in these more than in student teaching supervisor ratings.

Entrance and high school achievement data (Items 10, 11, 12 and 13) are not closely related to the ratings but are more nearly related to faculty than to field ratings. Surprisingly, they bear a somewhat closer relation to ratings made by academic faculty. The relationships are highest for the (ACE) Linguistic score and for the high school G. P. A.

Later Field Evaluation

Four years later, near the end of a fifth teaching year, a second field evaluation was obtained. Principals rated teachers on traits such as "appearance" and "industry" as well as in general on "personal" and "professional" qualities. Table II, Part 1, reports trait inter-relationships on the principals' evaluation. Some of the traits seem comparatively independent. The concomitant variation of "health" with "industry," "health" with "intelligence" and "health" with "professional knowledge" and other relationships seem consistent with what is known concerning these variables in other populations. It may be concluded that the data came from relatively careful rating.

Table II, Part 2, reports the relation between sub-category ratings and major categories. For example, it appears that "health" plays a minor part in the "professional" ranking but a major part in the "personal" one. "Social attitudes" count equally in either ranking and intelligence counts more in the "professional" rating. The reasonableness of the discovered inter-relationships lends strength to the supposition that principals can separately rate "personal" and "professional" qualities.

The major question is how does the later or "five-year" field rating agree with the first field (principals') rating and with original recommendations on file at the college. These data are reported in Table III. On "professional qualities" the later field evaluation agrees more with the original college evaluation than with the first field evaluation. Faculty ratings are better predictors for the fifth teaching year than for the first and evaluations from student teaching have gained most in predictive value. The ACE Linguistic score has also increased in value. There has been some decline in the value of placement office composite ratings.

While later field estimates moved toward agreement with college ratings on "professional" qualities, they diverge from those made on personal qualities. There is no evidence to indicate whether this results because: a) personalities change, b) because college does not elicit or evaluate adult personality manifestations, or c) because the first field ratings over-depend on personality. The study provides no explanation either for the fact that ratings from the second student teaching tend to be less predictive than ratings from the first.

Summary

Teaching competency as it will be judged by the principal can be predicted with roughly the accuracy that a final examination score may be predicted from a mid-term. The level of efficiency would probably reduce in a large depersonalized program or under severe departmentalization where academic faculty assumed a less responsible relation with teacher education.

The quality of teaching beyond the crucial minimum of failure is related to ability, college entrance test scores and academic success. It is at least as closely related to records of social participation and attitude. Some other studies find personality data quite superior to academic data or faculty ratings in predicting teacher performance.[1]

It appears that a multi-factor criterion is essential for the selection of teachers and it further appears possible and useful to distinguish between "personal" and "professional" fitness.

Selection and prediction methods are currently naive and untested. Are there specified traits which are either necessary or sufficient to predict teaching failure? At what points or in what ranges are they so significant? Are there alternative or contingent conditions which result in success? What are the qualities and the ranges of qualities that contribute to success? What are the ranges of teaching efficiency in which these operate? Better teacher selection depends at least in part on discovering more definitive criteria. It might be hoped that some agency or foundation would undertake an extensive and sophisticated study of this matter.

This study uncovers a significant relationship between teaching success and an early measure of democratic attitudes. These attitudes may foretell a capacity for empathy, a willingness to act in the interests of children or ability in managing subject matter. The stability, the cause and the level of this relationship is typical of problems which should be explored if teacher selection and placement are to be improved.

[1] David L. Cole, "The Prediction of Teaching Performance," The Journal of Educational Research, LIV, No. 9, May 1961.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

A STUDY OF THE CORRELATION BETWEEN AWARENESS OF STRUCTURAL RELATIONSHIPS IN ENGLISH AND ABILITY IN READING COMPREHENSION*

ROY C. O'DONNELL
Mount Olive Junior College
North Carolina

IT HAS OFTEN been asserted that lack of grammatical knowledge accounts for a great deal of incompetence in reading, but the various studies investigating the correlation of knowledge of grammar and ability in reading have offered no conclusive objective evidence to support the assertion. In fact, most of the studies that have been made indicate that the two are not highly correlated. It should be pointed out, however, that those who have investigated the relationship in question have not made a clear distinction between awareness of grammatical structure and knowledge of grammatical rules and terminology. They have generally based their findings on measurements of the degree of mastery of the conventional English grammar course, which, in the opinion of many modern grammarians, does not adequately represent the essentials of grammatical structure.

The Problem

The present study was designed to investigate the relationship between ability in reading comprehension and knowledge of grammatical structure. The specific question with which the study is concerned is whether there is a higher correlation between reading comprehension and awareness of structural relationships of words in sentences than there is between reading comprehension and ability to verbalize grammatical rules and terminology.

The investigation was limited to relationships that could be measured by written tests. Furthermore, it was concerned with knowledge and abilities acquired prior to the investigation, and no experimental teaching to compare relative effectiveness of different instructional approaches was involved.

The investigation of the problem necessitated the following steps: 1) basic structural relationships of words in English sentences were identified by examining textbooks based on structural linguistics; 2) a test was constructed to measure ability to recognize structural relationships; 3) the correlation of scores on the constructed test and scores on a standardized reading comprehension test was computed; 4) the correlation between scores on a standardized grammar test and reading test scores was computed; 5) the correlation of reading comprehension test scores and structure test scores was compared with the correlation of reading comprehension and grammar test scores; and 6) intelligence test scores were used to compute coefficients showing the correlation of the various tests with intelligence.

The Subjects

Arrangements were made in the spring of 1960 to give the reading test, the grammar test, and the structure test to the senior class at Bowling Green High School, Bowling Green, Kentucky.

The class was composed of 115 members — 64 boys and 51 girls. The Otis Quick-Scoring Mental Ability Test had been given to 106 of the 115 members of the class near the end of the preceding semester. At the time the Otis test was administered, the ages of those taking the test ranged from 15 years, 8 months, to 19 years, five months, with a median age of 17 years, eight months. The IQ score range for the class was from 80 to 129, or from the fifth percentile to the ninety-ninth percencentile, on the basis of data supplied by the test publisher. The mean IQ score was 106, corre-

* This report is based on a doctoral dissertation completed at Peabody College in January, 1961.

sponding to the seventieth percentile; the median IQ score was also 106.

The Procedure

Basic structural relationships of words in sentences were identified by examining four textbooks (Francis, Hill, Sledd, and Whitehall) incorporating principles of structural linguistics, and a test consisting of fifty items of the three-option multiple-response type was constructed to measure ability to recognize these relationships. In order to exclude as much lexical meaning as possible from the test items, nonsense words were used in the option sentences; however, the usual structural signals of word order, inflectional and derivational affixes, and English function words were provided.

The constructed test included six items designed to measure ability to recognize the relationship between subject and predicate, ten items to measure recognition of the relationship between verb and complement, four items to measure recognition of the relationship between coordinate elements, twenty-four items to measure recognition of relationships of the various types of modifiers and elements modified, and six items to measure recognition of the relationship between elements involved in cross-reference. The number of items of each type was decided upon after examination of the Stormzand and O'Shea study of the relative frequency of occurrence of various constructions in written English.

It was not feasible to establish validity of the test by an objective criterion, but every effort was made to construct the test so that correct responses would depend on ability to recognize the relationships involved. The reliability of the test was established by standard statistical procedures, and after preliminary testing, item analysis, and revision, the following statistical data were obtained: mean, 24.94; mode, 16; median, 24; standard deviation, 8.72; split-half reliability (Spearman-Brown formula), .88; inter-item consistency (Kuder-Richardson formula), .86.

In addition to the "Test of Recognition of Structural Relationships in English," constructed by the investigator, Test C 1: Reading Comprehension, Form Z, published by the Cooperative Test Division of the Educational Testing Service, and the Iowa Grammar Information Test, Form A, published by the University of Iowa, were administered to the subjects of the study. The number of students taking the grammar test was 113; the number taking the reading test was 108; and the number taking the structure test was 106. The number of students taking all three of the tests was 101, and Otis IQ scores were available for ninety-two of the 101 students.

The correlation coefficients of primary interest were those showing 1) the relationship between scores on the grammar test and scores on the reading test and 2) the relationship between scores on the structure test and scores on the reading test. Scores on the intelligence test were used to compute coefficients showing the relationship of the scores on the other tests to intelligence test scores; other coefficients were computed to show how the tests were related to one another.

The Pearson product-moment formula was used in computing the various correlations. Since only raw scores were obtained on the structure test, raw scores were used on the grammar test; scaled scores were used on the other tests. Correlation coefficients were computed using the separate level of comprehension and vocabulary scores on the reading tests; speed scores were considered irrelevant to the investigation. The .05 level was used in all tests of significance.

The Findings

The data showing the correlations of the various tests used in the study are presented in Table I and interpreted in the following paragraphs.

The correlation between structure test scores and Otis IQ scores was interpreted as being sufficiently low to invalidate any argument that the structure test is merely an intelligence test. The correlation of the Otis test and the grammar test was not found to be significantly higher than that of the Otis and structure tests. The correlation of the Otis and level of comprehension scores and of the Otis and vocabulary scores is in keeping with the various studies indicating a relatively high correlation between intelligence and reading ability.

The relatively high correlation of the grammar and structure test scores suggests that a common factor was being measured in both tests. The high correlation indicates that mastery of the terminology and rules of conventional grammar tends to be accompanied by ability to recognize basic structural relationships of words in sentences, but it does not necessarily imply that learning terminology and rules always results in ability to recognize structural relationships. It is possible that ability to recognize structural relationships enables the student to learn grammatical rules and terminology more easily.

The correlation of the grammar test scores and level of comprehension scores is not significantly different from the correlation of structure test scores and level of comprehension scores. This comparison does not support the hypothesis that there is a higher correlation between reading comprehension and ability to recognize structural relationships than there is between reading comprehension and knowledge of conventional grammar. In comparing the various correlations presented in Table I, however, we find that grammar test scores are more highly correlated with vocabulary

TABLE I
CORRELATIONS OF TEST C 1:
READING COMPREHENSION, THE IOWA GRAMMAR INFORMATION TEST, THE "TEST OF RECOGNITION OF STRUCTURAL RELATIONSHIPS IN ENGLISH," AND THE OTIS QUICK-SCORING MENTAL ABILITY TEST WITH ONE ANOTHER

	Reading Level	Reading Vocabulary	Iowa Grammar	Structure	Otis M. A.
Reading Level	--	.76	.46	.44	.65
Reading Vocabulary	--	--	.90	.46	.75
Iowa Grammar	--	--	--	.75	.57
Structure	--	--	--	--	.48

scores than are structure test scores. We also find that vocabulary and level of comprehension scores are relatively highly correlated. These data suggest that knowledge of vocabulary is an important factor in both the level of comprehension and grammar test scores but is of lesser importance in the structure test scores.

To obtain statistical evidence of the extent to which knowledge of vocabulary affects the correlation of level of comprehension and grammar test scores and level of comprehension and structure test scores, the partial correlation formula was used to partial out knowledge of vocabulary. With knowledge of vocabulary held constant, the correlation of the level of comprehension and structure test scores was .15 and the correlation of the level of comprehension and grammar test scores was -.79. Thus, with knowledge of vocabulary held constant, the correlation of the level of comprehension and grammar test scores is noticeably lower than that of level of comprehension and structure test scores.

The Conclusions

On the basis of the findings of the study, the following conclusions were drawn:

1. Mastery of the content of the conventional English grammar course tends to be accompanied by ability to recognize basic structural relationships of words in English sentences, to the degree that these abilities are indicated by scores on the tests used.
2. With knowledge of vocabulary accounted for by a partial correlation formula, the correlation between ability to comprehend written English and ability to recognize structural relationships of words in sentences is higher than the correlation between ability to comprehend written English and ability to verbalize knowledge of grammatical terminology and rules.
3. The correlation between awareness of structural relationships of words in sentences and ability in reading comprehension as indicated in this study is not sufficiently high to give conclusive evidence to support the teaching of linguistic structure as a major means of developing reading comprehension. Since there are several distinct factors involved in reading, however, it does not necessarily follow that ability to recognize structural relationships is of no importance in reading comprehension.

Recommendations for Further Study

Related problems that may merit further investigation are suggested as follows:

1. Since the present study is concerned with concepts of English structure acquired prior to the time of testing rather than the mode whereby the concepts may be acquired, it does not provide conclusive data proving that study of traditional grammar is the most effective means of forming these concepts. It would be of some value to know whether mastery of traditional grammar necessarily results in ability to recognize structural relationships or, conversely, whether a previously acquired ability to recognize structural relationships contributes to mastery of traditional grammar. If awareness of structural relationships is brought about by instruction, it would be useful to know whether instruction in structural linguistics is more, or less, effective than instruction in conventional grammar. A study of this kind might involve experimental teaching and testing on the elementary school level.
2. Since the "Test of Recognition of Structural Relationships in English" is constructed to measure ability to recognize structural relationships in English without direct application of knowledge of grammatical terminology, it appears that the test could be adapted to investigate the levels of maturity at which children develop awareness of the more complex grammatical relationships.
3. The structure test is dependent to a lesser degree than conventional grammar tests on knowledge of vocabulary, and thus it appears to more nearly isolate ability to recognize structural principles than do other tests. Therefore, it seems that the test could be used in research to discover more exactly the relative importance of the various factors in reading ability.
4. The present study is concerned only with the correlation of knowledge of grammatical structure and reading ability, but the structure test could be used to investigate the correlation of knowledge of structure and writing ability.

REFERENCES

1. Center, Stella S. The Art of Book Reading (New York: Charles Scribner's Sons, 1952).
2. Francis, W. Nelson. The Structure of American English (New York: The Ronald Press Company, 1958).
3. Hill, Archibald A. Introduction to Linguistic Structures (New York: Harcourt, Brace and Company, 1958).
4. Sledd, James. A Short Introduction to English Grammar (Chicago: Scott, Foresman and Company, 1959).
5. Stormzand, Martin J., and O'Shea, M. V. How Much English Grammar? (Baltimore: Warwick and York, Inc., 1924).
6. Whitehall, Harold. Structural Essentials of English (New York: Harcourt, Brace and Company, 1958).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

THE KINESTHETIC METHOD IN REMEDIAL READING

WILLIAM OFMAN, MORTON SHAEVITZ
University of California, Los Angeles

THE RETARDED READER, the person of normal intelligence who experiences difficulty in learning to read, has been a source of bewilderment and interest for some time (2, 4, 6). Recently, many institutions have been established which specialize in the treatment of persons with reading difficulties. The interest of workers in such centers has led to much theorizing about the causes and treatment of these people, and many hypotheses have been suggested to account for the phenomenon of an intelligent adult struggling with desperate intent to untangle a sentence which might be found in a fourth-grade reader (2: Ch.4, 5, 6; 6: Ch. 9, 10).

Despite this widespread interest, a review of the literature reveals a dearth of work dealing with the isolation of the pertinent variables responsible for the success or failure of any particular remedial method. In studies that have been reported, there is, as Gray points out (5, 1957: 401), in many instances an absence of statistical analysis, weak or almost no design, conclusions which are unwarranted, and a general failure to control for intervening variables.

One of the methods used in remedial work with retarded readers is described by Grace Fernald in her book (4), and sometimes referred to as the "Kinesthetic Method". Fernald defines the retarded reader as being one or more years retarded in reading skills, as measured by standardized tests. In general, she holds that persons in such circumstances have not been able to respond to the customary reading approach used in the public schools because they are "kinesthetic learners". Having postulated three types of imagery available to persons in the learning of reading: visual imagery, auditory imagery and kinesthetic imagery, Fernald assumed that children who could not learn to read in the usual visual-auditory manner are those who are so organized physiologically that they must utilize the kinesthetic imagery dimension. It can be shown that children who have made attempts at reading and failed have, in their efforts, utilized some kinesthetic cues. Observations by the authors have substantiated the commom observation that children move their eyes and their heads in reading as they follow the shapes of letters and words. Fernald, it appears, means something more by her method than the utilization of these kinesthetic cues.

As dictated by the Fernald method, a person is taught to read by responding to a felt need for a word in a story writing attempt. He asks the teacher for the word he needs but does not know how to write. The teacher writes the word on a slip of paper, pronouncing the word as it is written. The student is then asked to trace the word with his finger and pronounce it at the same time. The student traces the word as many times as he needs to before he can write and read it on his own. He uses the word in his story and may refer to it at any future time.

The authors know of cases where, as an adjunct to the Fernald method, and stemming from a belief in the crucial role played by the tactual-kinesthetic cues, words were cut out of sandpaper and finger-traced by students, thus increasing the tactual components involved.

The experimental evidence evaluating the effects of the addition of motor-tactual cues is sparse, though generally favorable. Early studies by Berman (1) indicated that the tracing of syllables on cards by his subjects did not aid in the learning of nonsense syllables, but in a later experiment he did find that the addition of motor cues helped in the learning of geometric figures. Pulliam (7) in attempting to evaluate the effectiveness of the addition of a tracing method in vocabulary building, had a group of retarded readers trace words on indented cardboard. His results indicated that the tracing method facilitated the acquisition of vocabulary in retarded readers.

In a recent and pertinent study by Roberts and Coleman (8), several of Fernald's (4) assumptions relative to the kinesthetic method were tested.

One of their hypotheses was that learning to read would be more efficient in retarded readers if kinesthetic elements were added to the learning-to-read process. The results of their experiment indicate that retarded readers were significantly better able to learn new materials if kinesthetic components were added, while the addition of kinesthetic elements did not aid normal readers. The authors conclude that their findings provide justification for the use of tracing methods in remedial reading programs.

It is submitted, however, that the essential variable of kinesthesis was not adequately isolated, and pertinent attention cues were not controlled, thus making unequivocal results impossible.

It was Fernald's point that children improved dramatically when they were taught to use a technique that added tactual-kinesthetic components to the reading process. She wrote: "Our study seems to show that normal perception, retention and memory for these same visual symbols could be developed in individuals whose failure seems to be due to inability to learn through normal visual and auditory channels if tactual and kinesthetic experiences were involved in the learning process . . . In all our cases of total disability . . . we have found that the individual learns rapidly and easily by the kinesthetic method . . . In other words, he actually needs to form the word with his hand and vocalize it during the initial learning process" (4:167).

The problem considered in this investigation was whether or not it is the tactual-kinesthetic variable that is the pertinent one in the learning of reading by retarded readers, and whether the reported improvement in reading ability by persons who could not learn to read by visual and auditory methods is indeed a function of the Fernald tracing method, and of the addition of tactual-kinesthetic imagery.

Method

The hypothesis tested was that the increase in reading ability by retarded readers is a function of the addition of tactual-kinesthetic cues to the learning process.

The control and isolation of the tactual-kinesthetic variable was accomplished by a device that allowed E to trace a word with a point of light from behind a screen. S, facing the screen, traced the word by following the moving light with his eyes. In effect, S saw a nonsense syllable resting on a screen of translucent glass, and a point of light which followed the inked outline of the syllable. In this manner, S had to attend to the word, and, depending on the group to which he belonged, trace it either with his finger or with his eyes.

The stimulus words were ten nonsense syllables of about seven percent association value drawn from Hilgard (Stevens, 9 : 544) and inked upon a slide in cursive writing.

Ss were 30 full-time male students enrolled at the University of California at Los Angeles Clinic School, a school established to help persons with reading difficulties by utilizing the Fernald method. Ss were randomly chosen from the student population, and assigned at random to one of three experimental conditions. The three groups of ten subjects each did not differ significantly in intelligence, age, reading ability, and length of training at the Clinic School. The mean age of the subjects was 13.5 years with a standard deviation of 3.16. In the experiment, Ss were re-pooled and the order in which they were run was randomized to eliminate any systematic biases.

The three treatments were:

Condition A: Eye Tracing. Ss traced the word presented them by visually following a moving point of light as it followed the outline of the syllable.

Condition B: Finger-Contact Tracing. Ss placed a finger upon the slide, were cautioned to press down, and traced the word with the finger, pronouncing the word as E had done.

Condition C: Simple Reading. Ss would merely read the word after E had pronounced it.

In each condition, the time for each segment of the procedure was equal, the method of presentation, and the number of exposures were identical.

Perhaps a description of the instructions and the procedure followed in the Eye Tracing condition will be clarifying. Ss were told: "As you look into the window, you will see a three-letter word which has no meaning. I will say the word for you. Then I want you to say the word just as I did. Then I want you to follow the light with your eyes, and as you trace the word with your eyes, I want you to say the word again." E then placed a slide into the window, and allowed a three-second interval. E pronounced the word and then S was instructed to trace and pronounce the word synchronously. Fifteen seconds of exposure time were allowed for each syllable under all conditions. The list of ten syllables was gone over three times with each S. After the third exposure to the list, S was tested by asking him to write the word as E pronounced it. This was the test, and the criterion score was the number of correct reproductions.

Results

An analysis of variance of the criterion scores for the three different conditions revealed no absolute difference between the mean scores for the eye-tracing and the finger-tracing conditions. The mean number of correctly reproduced words in the eye-tracing condition was 5.4, with a standard deviation of 2.05. The mean number of correctly

reproduced words in the finger-tracing condition was also 5.4, with a standard deviation of 1.80. In simple reading, the mean number of words correctly reproduced was 3.0, with a standard deviation of 1.89. A simple approximate test of homogeneity of variance indicated that homogeneity prevailed (3: 328), and an inspection of criterion measures suggested that the condition of normality was satisfied. An F of 4.33 indicated that the difference among the groups was significant beyond the .05 level of confidence, and a t-test computed between groups A or B, and C was 2.57 and 2.75 respectively and significant beyond the .05 level.

The null hypothesis was rejected. The data suggested that the eye-tracing and finger-tracing methods were both more effective than simple reading, but that finger-tracing was not more effective than eye-tracing in the acquisition of new material by a random sample of retarded readers.

Discussion

The pertinent variable in this experiment was the tactual-kinesthetic variable. With most relevant conditions held constant, the effectiveness of the tactual-kinesthetic method was tested. The results indicated that eye-tracing was equally effective as finger-tracing. It was concluded that the tactual-kinesthetic variable was not the relevant one in helping a population of non-readers (a population of which the random sample in this experiment may be representative) to learn new words.

It might be valuable to note that in the experiment the "cards were stacked" against the hypothesis. It was expected that Ss would perform better under a condition with which they had previous experience. All Ss were students in the Clinic School and had much training in the finger-contact method. The results indicate, however, that the cues gained from the addition of tactual stimulation did not increase the learning ability of this population. The fact that this group had had experience with the tracing method strengthens the conclusion that tactual cues, qua tactual cues, may not be seminal in the learning of reading by retarded readers.

In order to be able to generalize from this study to retarded readers generally, it would, of course, be necessary to replicate the procedure in various settings.

It is possible to speculate regarding other variables which may account for the obtained equivalent results between the finger-tracing and eye-tracing methods and the reported success of the finger-tracing method generally. The finger-tracing method demands that S approach the reading task in a special manner. S must come in contact with another person who gives him the word, he must, as he traces the word with his finger, attend to his task (i.e., look at the word), he focuses on the word from left to right, he utilizes kinesthetic, tactual, visual and auditory cues as he traces the word. The salient feature in the finger- and eye-tracing methods seems to be to compel S to attend to the task at hand, as opposed to the simple reading technique where attention is not so closely controlled. One might consider it possible that the success of the kinesthetic method as reported in the literature is a function of its ability as an attention-holding device.

In regard to certain practical questions which might be raised, the findings both support and impugn certain aspects of the tactual-kinesthetic method. On the one hand, it appears that the tactual-kinesthetic variable may not be pivotal insofar as the increase in reading ability by retarded readers is concerned. On the other hand, if the important relationship is between attention to the learning task and the increase in performance, then the tracing method seems to be quite economical, and eminently suited in this regard.

Summary

The relationship between the learning of reading by retarded readers and the addition of tactual-kinesthetic cues in the learning process was tested. Thirty Ss, all retarded in their reading to a comparable degree, enrolled in a remedial reading program which utilized the Fernald method, were divided into three groups of ten, comparable on some pertinent variables. They were subjected to three different learning conditions utilizing varying amounts of kinesthetic cues. Ss learned low association nonsense syllables by utilizing either an eye-tracing, finger-tracing, or simple reading technique. All Ss had been previously trained in the finger-tracing method.

The results revealed no difference between the tactual-kinesthetic and the eye-tracing method, but that both of these methods were superior to the simple reading method in the learning of words as tested by a written recall task. It was speculated that the variable of which the learning to read by retarded readers is a function may be the attention variable, rather than the previously supposed tactual-kinesthetic one.

REFERENCES

1. Berman, A. "The Influence of the Kinesthetic Factor in the Perception of Symbols in Partial Reading Disability," Journal of Educational Psychology, XXX (1939), pp. 187-198.
2. Bond, Guy L., and Tinker, Miles A. Reading Difficulties, Their Diagnosis and Correction (New York: Appleton-Century-Crofts, 1957).
3. Edwards, Allen L. Statistical Methods for the Behavioral Sciences (New York: Rinehart and Company, 1954).

4. Fernald, Grace M. Remedial Techniques in the Basic School Subjects (New York: McGraw-Hill Book Co., 1943).
5. Gray, William S. "A Summary of Reading Investigations," Journal of Educational Research, XXXIX-LI (1946-1957).
6. Harris, Albert J. How To Increase Reading Ability (New York: Longman's, Green, 1956), third edition.
7. Pulliam, R. A., et al. "Indented Word Cards as a Sensori-Motor Aid in Vocabulary Development," Peabody Journal of Education, XXIII (1945), pp. 38-42.
8. Roberts, Richard W., and Coleman, James C. "An Investigation of the Role of Visual and Kinesthetic Factors in Reading Failure," Journal of Educational Research, LI (1958), pp. 445-451.
9. Stevens, S. Handbook of Experimental Psychology (New York: Wiley, 1951).

TABLE I

SUMMARY OF DATA

Groups	Eye-tracing	Finger-tracing	Reading
Mean number of words learned	5.4	5.4	3.0
Standard deviation	2.05	1.80	1.89

Overall F with 2, 23df is 4.33; p is less than .05

t test between groups A or B, and C is 2.57 and 2.75 respectively, and significant beyond .05 level.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

IMPROVEMENT OF CRITICAL THINKING IN RELATION TO OPEN-CLOSED BELIEF SYSTEMS

C. GRATTON KEMP
Ohio State University

THE QUALITY of our thinking will determine our existence as a free people. In the past this ability was required of the few, now it is recognized as a necessity for the many. So complex and rapidly changing is our society that each citizen must become adequate in the exercise of the higher processes of thought. Improvement in critical thinking is an urgent necessity.

This is apparent in all disciplines by the growing emphasis placed upon training in critical thinking. The need is focused by educators who emphasize that more assistance be given to students to "develop problem-solving methods which will yield more complete and adequate solutions in a wide range of problem situations." 1* Progress has been made in at least three directions: a recognition of the need for improvement; a clarification of the meaning of the term, critical thinking; and analysis of the deterrents to better performance. In the definition of the concept of critical thinking reference has been made to five abilities:

1. The ability to define a problem.
2. The ability to select pertinent information for the solution of a problem.
3. The ability to recognize stated and unstated assumptions.
4. The ability to formulate and select relevant and promising hypotheses.
5. The ability to draw conclusions validly and to judge the validity of inference. [2]

In the analyses of the deterrents to better performance it has been concluded that:

1. They tend to avoid real problem solving.
2. They apply only a limited stock of techniques to solve them.
3. They are satisfied with a partial solution.
4. They change the problem completely.
5. They escape from it entirely. [3]

These behaviors suggest the influence of emotional factors on critical thinking. The research of Else Frenkel-Brunswik,[4] Postman,[5] Allport,[6] Maslow,[7] Rokeach,[8] and others provide confirmation of the influence of the emotions on the cognitive spheres (thinking, perception, and memory). In a study[9] of the influence of dogmatism on critical thinking, it was found that high dogmatism or closed-mindedness decreased efficiency. It was concluded that this less efficient performance of the high dogmatic was due to the following factors:

1. Difficulty in tolerating ambiguities which leads to a "closure" before full consideration is given to each piece of contributing evidence.
2. A perceptual distortion of facts resulting in a decision which does not encompass all elements of the problem.
3. Lack of recognition or rejection of significant parts or of the whole problem in order to accommodate it into the preformed value pattern, resulting in a poor or incorrect solution. [10]

On the other hand, "the more open-minded perceptively examine all aspects of the experience, try to clarify the ambiguity, and strive to see the relationship among parts."[11]

The contrast in the performance of those with an open belief system when compared with those having a closed belief system led to the assumption that improvement in critical thinking might also be influenced by the factor of dogmatism.

* Footnotes will be found at the end of the article.

Purpose and Hypothesis of the Study

The purpose of the study was to compare the improvement in critical thinking of those low in dogmatism with those who were high.

It was hypothesized that those low in dogmatism would show greater improvement in critical thinking than those who were high.

Procedures

A total of 80 freshman students, 40 in the control group and 40 in the experimental group, participated in the study. The two groups were matched in intelligence and in degree of open- and closed-mindedness.

The Dogmatism Scale Form E[12] and the Otis Test of Mental Ability Form A were administered to the total sample of 80 students. On the basis of these criteria two comparable groups of 40 each were established. Both groups also took the Watson-Glaser Critical Thinking Appraisal.

The Control Group and Experimental Group were taught by the same instructor. The Control Group received no special help in solving critical thinking problems.

The Experimental Group was divided into five subgroups, of eight each, with four high and four low in dogmatism in each subgroup. Each of these subgroups participated in 10 one-hour meetings. Each meeting was used in solving critical thinking problems. Reasons for correct and incorrect conclusions were discussed. These were recorded and those which were general to the five groups are listed below as factors which affected efficiency.

Factors Which Increased Efficiency

1. Synthesis and evaluation of process relationships.
2. Evaluation of conclusions for inclusiveness and basicity.
3. A flexible approach.
4. Adherence to the conditions given.
5. Analysis of data provided.

Factors Which Decreased Efficiency

1. Ignore, distort or omit some of the given data.
2. Include additional words and/or ideas.
3. A rigid, often nonadaptive approach.
4. Failure to synthesize or a poor quality of synthesis.
5. Failure to evaluate conclusion or incompleteness of evaluation.

For each problem the student stated what he considered to be the factor or factors which led to a correct or incorrect conclusion. These factors are stated above in the rank order established on the frequency of occurrence indicated by the student with reference to those which either increased or decreased efficiency.

The critical thinking problems were selected from the following forms: Taxonomy of Educational Objectives; A Test of Problem Solving (High School Edition); and A Test of Critical Thinking Form G.

At the end of the study both groups were again administered the Watson-Glaser Critical Thinking Appraisal, and the results compared.

Analysis of the Data

The use of the Wilcoxon Matched-Pairs Signed-Ranks Test[13] gave the following results:

At the Beginning of the Study:

1. There was no significant difference in critical thinking between the Control Group and the Experimental Group.
2. In each of the Control and Experimental Groups, the "Low" did significantly better than the "High" at the five percent level of significance.

At the Completion of the Study:

1. The Experimental Group did significantly better in critical thinking than the Control Group at the five percent level.
2. In the Control Group, there was no significant improvement in the performance of either the "Low" or "High."
3. In the Experimental Group, the "High" improved but the improvement was not significant; the "Low" performed significantly better at the one percent level.

Conclusions

1. Improvement in critical thinking (as measured by the Watson-Glaser Critical Thinking Appraisal), is unlikely in the usual classroom situation.
2. Under favorable conditions those with open minds (as measured by the Rokeach Dogmatism Scale Form E), show greater improvement in critical thinking than those with closed minds.
3. The favorable conditions are permissive (safe) small group situations in which the usual threats are minimized and in which intensive attention given to the factors in critical thinking is accompanied by extensive practice.

Implications

1. Since dogmatism is learned, attention in the early life of the child should be given to the provision of a learning environment which encourages open-mindedness. This environment will provide safety, and encourage self-awareness and self-evaluation of internal and external stimuli.
2. Throughout the school life of the individual more opportunity should be provided for engaging in critical thinking.
3. More efficiency in critical thinking could be expected in those classroom environments which reduce the degree of threat, affording as much permissiveness (safety) as the students demonstrate they can use.
4. Emphasis should be placed upon helping the individual to understand those inefficiencies which are a result of dogmatism on both his attack on the problem and his performance in solving it.

FOOTNOTES

1. Benjamin S. Bloom, Editor, Taxonomy of Educational Objectives (New York: Longmans, Green and Co., 1956), p. 43.
2. Paul L. Dressel and Lewis B. Mayhew, General Education: Explorations in Evaluation (Washington, D.C.: American Council on Education, 1954), pp. 179-181.
3. Bloom, op. cit., pp. 42-43.
4. Else Frenkel-Brunswik, "Intolerance of Ambiguity as an Emotional Perceptual Personality Variable," Journal of Personality (1949), pp. 18, 108-143.
5. L. Postman, J. Bruner, and E. McGinis, "Personal Values as Selective Factors in Perception," Journal of Abnormal and Social Psychology (1948), pp. 43, 142-154.
6. Gordon W. Allport and L. Postman, The Psychology of Rumor (New York: Henry Holt and Co., 1947), chapter 6.
7. A. H. Maslow, Motivation and Personality (New York: Harper and Bros., 1954), p. 266.
8. Milton Rokeach, The Open and Closed Mind (New York: Basic Books, 1960), chapter 3.
9. C. Gratton Kemp, "Effect of Dogmatism on Critical Thinking," Journal of School Science and Mathematics (April 1960), p. 318.
10. Ibid.
11. Ibid., p. 315.
12. Rokeach, op. cit., p. 72 ff.
13. Sidney Siegel, Nonparametric Statistics (New York: McGraw-Hill Book. Co.), pp. 75-83.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, March 1963)

COLOR IN DISCRIMINATIVE LEARNING BY CHILDREN

J. J. STALEY, WILLARD H. NELSON, and F. J. KING
The Florida State University

DURING RECENT YEARS a number of studies of discriminative learning in children have demonstrated that elementary school pupils (kindergarten through grade five) discriminated less well between stimulus objects which differed only in hue than between those which differed only in shape or in degrees of brightness (2, 3, 4). Clifford and Calvin (4) reported that in a discrimination problem where color was the cue, 63% of all Ss failed to reach criterion of ten consecutive successful trials. In a similar problem where black and white were used instead of color, only 38% of the subjects failed. Brackbill and Jack (1) investigated reinforcement value on discriminative learning in children. Holding amount plus probability of occurrence of reinforcement constant, they found that children who were allowed to choose their reinforcers showed significantly less variability in reaching a learning criterion than did children who were not allowed such a choice. The present study was designed to investigate the effects of intelligence, sex, grade level, and choice vs. specified reinforcement on discriminative learning where color cues were used.

Method

Subjects. The subjects were 128 children from a public school in Tallahassee, Florida. Of this total, 64 were in the first grade and 64 in the third grade. Each group was made up of equal numbers of boys and girls.

Apparatus. An exposure apparatus constructed of white cardboard was used to present the materials. Its central panel, 22 inches wide and 28 inches high, operated by E, was the exposure screen. The stimulus-cards, each 5 × 5 inches, colored Calamine Blue and Shamrock Green as determined from the color chart in the unabridged second edition of Webster's New International Dictionary, 1951, were behind the screen and in front of identical egg cups, under one of which was a ring. The cards, and the cups behind them, were 4 inches apart and even in the frontal plane.

Procedure. Ss in each grade were divided into two groups of 32, equal as to sex. The third grade groups were divided into subgroups of high intelligence (median IQ 127.5), and average intelligence (median 101.5). Each S was required to discriminate between blue and green cards. As each S sat at a table with the apparatus before him E operated the screen, arranging the cards, cups, and ring in their proper positions behind the screen for each presentation. Cards were presented in random order with the limitation that a card did not appear on the same side more than twice in succession. The reward for Ss in the first group in each grade was a balloon. Subjects in the second group could choose among the balloon, a piece of candy, and a piece of bubble gum.

The criterion of learning was five consecutive correct trials. The test was terminated at the first trial after criterion, or at the first error after the nineteenth trial if the criterion had not been reached. Ss not learning after twenty trials were given an arbitrary score of 25. Ss were requested to identify colors after experiment; all did so without error.

Instructions. Upon entering the testing room, S was seated at the exposure apparatus and given the following instructions: "Will you play a game with me? I have two cards. Behind each card is a cup. Under one of the cups will be a ring. Each time that you guess which cup the ring is under, you will get a balloon." (First groups.) The instructions for each of the second groups were the same except that, "... will get the prize that you choose."

Results

The effects of sex, grade level, and choice vs. specified reward were examined through analysis of variance. This analysis involved the total population, and no significant mean differences were found.

A second analysis was performed on the scores of the third graders to investigate effects of sex, choice vs. specified reward and intelligence level. The only significant F-ratio (.05 level) was found to exist in the interaction between sex and intelli-

gence level; Duncan's New Multiple Range Test (5) showed the differences to be between boys of different intelligence levels and between boys and girls of high intelligence. Mean scores to criterion for third graders were as follows: boys of high intelligence, 14.00; boys of average intelligence, 20.32; girls of high intelligence, 19.38; girls of average intelligence, 18.12.

Discussion

These findings tend to support those of Calvin, et al., that color cues do not elicit efficient learning in discrimination problems. Although analysis of the data revealed no explanation for this situation, it is of interest that every child who named a color in his first choice ("It is behind the blue (green) one.") reached criterion. This could have indicated a realization that color may function to provide a key to problems involving differentiation of two otherwise similar objects; the better performance of children with black and white discrimation choices might then be shown to result from the more ordinary and obvious contrast between black and white than between blue and green. Further research might well be carried out to discover whether the use of two of the primary colors would produce better results than does the use of one primary and one of its adjacent secondaries.

Since each of the subjects was able to identify each of the two colors used on the stimulus cards, it is obvious that simple inability to discriminate between them did not account for lack of ability to solve the problem. There was no indication in the above results, however, as to what might constitute a reasonable explanation of the relation of color to discrimination where discrimination is a means to an end rather than an end in itself.

The significant interaction of sex and level of intelligence is also impossible to explain on the basis of present results. The superior performance of boys of high intelligence over all other groups could possibly be related both to sex differences in certain mental abilities and also to cultural factors which tend to predispose boys more than girls to be interested in certain types of abstract, "rules-governed" puzzles and games. Further research would be necessary to validate this interaction and to examine its relationships to other factors.

Although these findings are unrelated to use of color for attractiveness in text books, charts, etc., there is some indication that use of color as a cue in problems prepared for children in the primary grades may not promote rapid and efficient learning. Readiness workbooks and first arithmetic and spelling books are examples of places where this kind of color use might prove to be ineffective.

Summary

Discriminative learning of 128 children, 32 boys and 32 girls from each of the first and third grades, was investigated with a problem involving color. No differences in performance due to grade or type of reinforcement were found. An interaction of sex and intelligence level was found to be significant at the .05 level.

REFERENCES

1. Brackbill, Yvonne and Jack, D. "Discriminative Learning in Children as a Function of Reinforcement Value," Child Development, 1958, XXIX, 185-90.
2. Calvin, A. D., and Clancy, J. J. and Fuller, J. B. "A Further Investigation of Various Stimulus-Objects in Discriminative Learning by Children," American Journal of Psychology, 1956, LXIX, 103-6.
3. Calvin, A. D. and Clifford, L. T. "The Relative Efficiency of Various Types of Stimulus-Objects in Discriminative Learning," American Journal of Psychology, 1956, LXIX, 647-49.
4. Clifford, L. T. and Calvin, A. D. "Effect of Age on Discriminative Learning of Color and Brightness by Children," American Journal of Psychology, 1958, LXXVII, 766-67.
5. Edwards, A. L. "Experimental Design in Psychological Research," (New York: Rinehart and Co., Inc., 1960) LXXVII, 136-37.

Special Section on

Measurements, Statistics, and Methods of Experimental Research

EDITED BY

RAYMOND O. COLLIER, Jr.
Professor, College of Education
The University of Minnesota

and

JULIAN C. STANLEY
Professor of Educational Psychology
The University of Wisconsin

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

RESPONSE SURFACE METHODOLOGY IN EDUCATION AND PSYCHOLOGY

DONALD L. MEYER
Syracuse University

THE PURPOSE of this paper is to call to the attention of experimenters in education and psychology the use of response surface methodology. Previous writers of response surface methods have been concerned with the statistical theory of these methods or in their application to industrial experiments.

The central problem is the estimation of a theoretical response surface or function, ϕ. This function is assumed to be dependent upon k continuous factors, X_i $(i = 1, 2, \ldots k)$, and to be representable by some finite degree polynomial.

That is, response surface methods are applicable when investigating a number of factors in order to assess their effect upon some response, and a functional relationship between the response and the factors is postulated. The concept of such a functional relationship is not new. Robinson proposed a two-dimensional relationship between "efficiency of recall" and "degree of similarity between interpolated activity and original memorization" in his study of the phenomenon of retroactive inhibition. This was later extended to the "Osgood transfer surface," which was a three-dimensional surface relating recall with "response similarity" and "stimulus similarity". (See, for example, McGeoch and Irion, 9).

There are at least three phases in a factorial-type investigation. First, the factors thought to be important and the particular levels of these factors to be used in the experimental program are selected. Second, the experiment is performed and the factors are determined to be statistically significant or not. Third, the functional relationship currently thought to describe the response is either upheld or modified. Oftentimes, it appears that many factorial programs are ended with the second phase. This may not always be undesirable, but one should be clear about what is meant by a "significant factor. Significance is a function of the spacing between the levels of the factor. For example, suppose that one of the factors was "time" and that the particular levels used were two hours and three hours. We may find that the difference in responses at these levels are not significant, but this does not mean the factor is not significant. If we used a wider spacing between the two times, we might then find that time would be a significant factor.

Response surface methodology is focused not only on the factors but primarily on the existing functional relationship. It would seem that progress in developing models and theory describing such relationships would be enhanced by just such a concern.

Definitions and Notation

Assuming the surface is approximated by a polynomial, the equation of the surface can be represented by

$$(1)\ n = B_0X_0 + B_1X_1 + B_2X_2 + \ldots + B_kX_k + B_{11}X_1^2 + \ldots + B_{kk}X_k^2 + B_{12}X_1X_2 + B_{13}X_1X_3 + \ldots + B_{k-1}, X_{k-1}X_k + B_{111}X_1^3 + \text{etc.}$$

where n is the theoretical response. N experiments are performed at N points in the k-dimensional factor space and responses, Y_u $(u=1, 2, \ldots N)$, are observed. The observed responses are assumed to be

$$(2)\ Y = n + e,$$

where $\underline{e}$ is a vector of residuals assumed to be normally and independently distributed.

Using ordinary least-squares theory, the vector of estimates, B, is given by

$$(3)\ \hat{B} = (X'X)^{-1}X'Y, \quad \text{where}$$

$$\hat{B}' = (b_0, b_1, b_2, \ldots, b_k, b_{11}, \ldots, b_{kk}, b_{12}, \ldots)$$

$$Y' = (y_1, y_2, \ldots, y_N)$$

$$X = \begin{bmatrix} 1, & X_{11}, & X_{21}, & \cdots & X_{k1}, & X_{21}^2, & X_{21}^2, & \cdots & X_{k1}^2, & X_{11}X_{21}, & \cdots \\ 1, & X_{12}, & X_{22}, & \cdots & X_{k2}, & X_{12}^2, & X_{22}^2, & \cdots & X_{k2}^2, & X_{12}X_{22}, & \cdots \\ \vdots & & & & & & & & & & \\ 1, & X_{1N}, & X_{2N}, & & X_{kN}, & X_{1N}^2, & X_{2N}^2, & \cdots & X_{kN}^2, & X_{1N}X_{2N}, & \end{bmatrix}$$

The X matrix is seen to be an N by L matrix whose u^{th} row contains the values of the L variables $X_0, X_1, \ldots, X_k, X_1^2$ etc. in (1), for the u^{th} experimental point. L is the number of terms in the polynomial and given by

(4) $L = \binom{k+d}{d}$,

where d is the degree of the polynomial used to represent the surface. A design used to estimate the coefficients in a polynomial of degree, d, is said to be a design of order d.

It can be noted that the symbolism developed above is essentially the same as that encountered in a multiple regression problem. There, the X matrix would be the scores, say, on L variables for N students, and the X'X matrix would be the sums of squares and cross-products matrix. The difference between a multiple regression problem and the response surface problem is that in the former case the values of the variables are found by testing the N individuals or otherwise examining their status, while in the latter case, the variables are experimental factors whose values are predetermined by design.

The design problem is, then, the selection of N values of the experimental factors (or the selection of N points in the k-factor space) in order to estimate the function ϕ. The classical factorial designs are available for this purpose, but there is no reason why designs should be limited to this class when the primary interest of the experimenter is focused on the response surface as a function of continuous factors. For example, the 3^2 factorial for two continuous factors requires experimentation at nine points in the two factor space as shown in figure 1 (where X_1 and X_2 are the two experimental factors.)

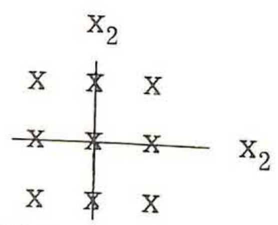


Figure 1: The 3^2 Factorial

This design has the property that the estimates of the factor effects are orthogonal, but when interest resides in the response function it may be more efficient to place eight of the points on a circle as shown in figure 2.

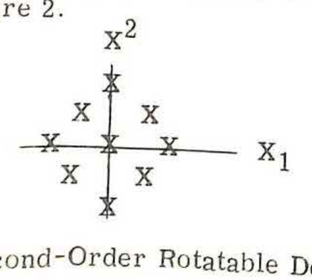


Figure 2: Second-Order Rotatable Design

Designs such as these are a subset of a class called "generalized factorials."

It is evident that there are infinitely many combinations of factor levels which could be used for a particular experimental program. A criterion for choosing a particular design was advanced by Box and Hunter (3) and is known as the criterion of rotatability. This criterion arises from consideration of the variance of an estimated response, y, at various points in the k-dimensional factor space. Each k-dimensional design of order d has a unique "variance function" which provides a measure of the precision of the estimated response at any point in the factor space.

According to the rotatability criterion, one would choose a design such that for any two points in the factor space which are the same distance, P, from the center of the design (i. e., $X_1 = X_2 = \ldots = X_k = 0$) the two values of the variance function for the two points are equal. If this criterion is met the design is said to be rotatable. That is, the contours of the variance function at distances, P_1, P_2, etc. are circles, spheres, or hyperspheres (depending on k). Rotatable designs can also be rotated to a different orientation relative to the axes of the factors without changing the rotatability property of the variance function. The 3^2 factorial design given above is not rotatable and upon rotation even loses the property of orthogonality.

A user of rotatable designs is expressing his ignorance about the response surface and its relative orientation by adopting a strategy of indifference as to the directions of interest.

Experimental Strategy and Design Requirements

An experimenter desiring to estimate a response surface must make several decisions such as selecting the particular factors to be investigated, the region of experimental interest, the scalings of the factors, and the degree of the model used to represent the surface. These decisions may be influenced by previous experimentation or by the experimenter's judgment. The region and scalings of the factors are chosen when the equations for transforming the experimental factors into standardized design variables are decided upon as shown below:

(5) $X_i = \dfrac{E_i - E_{io}}{S_i}$,

where E_{io} is some "middle" level of the i^{th} factor and S_i is the selected scaling.

Typically, a first-order model would be assumed until the data indicated that a higher-order model was necessary.

One would desire a design that was "sequential" in the sense that it could be augmented into a higher-order design and furthermore, that would al-

TABLE I

DATA FOR FIRST STAGE

Subject	X_1	X_2	X_3	X_4	E_1	E_2	E_3	E_4	Y(%)
1	1	1	1	1	4	20	9	4	27
2									40
3	1	-1	1	-1	4	10	9	2	50
4									56
5	1	-1	-1	1	4	10	3	4	85
6									100
7	1	1	-1	-1	4	20	3	2	70
8									111
9	0	0	0	0	3	15	6	3	83
10									44
11									33
12									60
13	-1	1	-1	1	2	20	3	4	75
14									140
15	-1	1	1	-1	2	20	9	2	67
16									0
17	-1	-1	1	1	2	10	0	4	80
18									100
19	-1	-1	-1	-1	2	10	3	2	120
20									113

TABLE II

ANALYSIS OF VARIANCE FOR FIRST STAGE

Source	d. f.	SS	MS	F
Due to regression	5	114814.706		
b_0	1	102961.250		
b_1	1	1172.994		1.81
b_2	1	1501.485		2.32
b_3	1	8788.875		13.57
b_4	1	390.102		.60
About regression	15	10220.294		
Error	11	7124.500	647.681	
Lack of fit	4	3095.794	773.948	1.19
Total	20	125035		

TABLE III

DATA FOR SECOND STAGE

Subject	X_1	X_2	X_3	X_4	E_1	E_2	E_3	E_4	Y(%)
21	1	1	1	-1	6	20	9	2	87
22									77
23	1	-1	1	1	6	10	9	4	56
24									67
25	1	1	-1	1	6	20	3	4	111
26									100
27	1	-1	-1	-1	6	10	3	2	108
28									50
29	0	0	0	0					80
30									83
31									100
32									67

low an estimation of the goodness of fit of the assumed model. One such design is the first-order triangle design shown in figure 3.

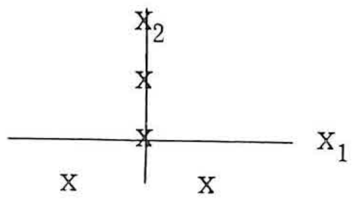


Figure 3: First-order Triangle Design.

With two observations at the center, the values of the design variables are

	Points				
	(1)	(2)	(3)	(4)	(5)
X_1:	0	0	0	$\sqrt{6}/2$	$-\sqrt{6}/2$
X_2:	0	0	$\sqrt{2}$	$-\sqrt{2}/2$	$-\sqrt{2}/2$

This design allows estimation of the coefficients in the first-order model:

(6) $y = b_0X_0 + b_1X_1 + b_2X_2$.

The analysis of variance for a single replicate would be as shown below.

Source	degrees of freedom
Due to regression	3
b_0	1
b_1	1
b_2	1
About regression	2
Error	1
Lack of fit	1
Total	5

It should be noted that the "lack of fit" mean square can be compared with the error mean square to give an indication of how adequate the assumed model is. In most applications several replicates of this basic design would probably be used in order to provide a larger number of degrees of freedom for error. For "r" replications the degrees of freedom for error would be 5r - 4 (i.e. minus three for the three coefficients and minus one for lack of fit).

This design can be augmented into a second-order design by adding points at

	Points				
	(6)	(7)	(8)	(9)	(10)
X_1:	0	0	0	$\sqrt{6}/2$	$-\sqrt{6}/2$
X_2:	0	0	$\sqrt{2}$	$\sqrt{2}/2$	$\sqrt{2}/2$

This forms the second-order "hexagon" design shown in figure 4.

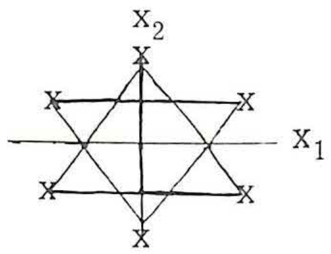


Figure 4: Second-Order Hexagon Design

The addition of two more points at the center insures that the two stages of experimentation form two orthogonal blocks. Consequently, the first five subjects could be rats, say from a single litter and the second five could be rats from a different litter. One degree of freedom would be allocated to "blocks". The fitted model would be

(7) $y = b_0X_0 + b_1X_1 + b_2X_2 + b_{11}X_1^2 + b_{22}X_2^2 + b_{12}X_1X_2$,

with the corresponding analysis of variance for a single replicate into six degrees of freedom for the six coefficients, three degrees of freedom for error, and one for lack of fit.

Several designs which can be utilized in this fashion are available. The two-level factorials and their fractional replicates, for example, can be used for first-order designs and can be augmented into the so-called "central composite" designs. These designs have been discussed in the articles given in the reference section.

An Illustration in a Human Learning Experiment

The experiment to be described is an illustration of the application of response surface methods. No claim is made as to the optimality of the particular experimental program, but it demonstrates the sequential nature of the methods.

The purpose of the experiment was to study the effect of four factors on the amount of retroactive inhibition induced in a typical "learning, learning a new task, recall" situation. The subjects were students in an introductory class in the psychology of learning at the University of Minnesota. The students were, for the most part, candidates for the masters degree in Education.

The experiment was performed in three stages. Within a particular stage, subjects were assigned to the treatments at random, but subjects were not assigned to the stages themselves at random. This limitation arose due to scheduling difficulties. Orthogonal blocking was planned but was not achieved due to change in the experimental region studied as will be seen below.

Experimental Procedure

Fifteen pairs of nonsense syllables were selected from the list of syllables given by Glaze (8). Each pair of syllables was typed on a three-by-five

TABLE IV

ANALYSIS OF VARIANCE FOR SECOND STAGE

Source	d.f.	SS	MS	F
Due to regression	5	122025.950		
b_0	1	116281.250		
b_1	1	885.270		2.48
b_2	1	162.690		.47
b_3	1	4693.620		13.64
b_4	1	33.120		.10
About regression	15	5771.050		
Error	11	3786.000	344.182	
Lack of fit	4	1985.050	496.263	1.44
Total	20	127797		

TABLE V

DATA FOR THIRD STAGE

Subject	X_1	X_2	X_3	E_1	E_2	E_3	Y(%)
33	1	1	0	8	30	6	50
34							79
35	1	0	1	8	15	9	75
36							62
37	1	0	-1	8	15	3	93
38							118
39	1	-1	0	8	0	6	100
40							63
41	0	1	1	6	30	9	87
42							67
43	0	1	-1	6	30	3	100
44							100
45	0	-1	1	6	0	9	67
46							93
47	0	-1	-1	6	0	3	92
48							100
49	-1	1	0	4	30	6	75
50							109
51	-1	0	1	4	15	9	38
52							73
53	-1	0	-1	4	15	3	92
54							91
55	-1	-1	0	4	0	6	54
56							67
57	0	0	0	6	15	6	86
58							92
59							92
60							86
61							92
62							77
63							87
64							100

inch white card, with the stimulus syllable on the left side and the response syllable on the right-hand half of the card. The cards were shuffled before every trial and presented to a particular subject (S) so that only the stimulus syllable was visible. S was told to pronounce the stimulus syllable as best he could. After three seconds he was shown the response syllable for three seconds. S was instructed to both pronounce and spell the response syllable. After all fifteen pairs had been given, S was given the same pairs again in a new random order with the instructions to pronounce the stimulus syllable and give the response syllable if he could. The response syllable was then exposed for three seconds whether S correctly anticipated or not.

After a particular pair had been given correctly a certain number of times (E_1), This particular pair was dropped from the list. S continued in this manner until all fifteen pairs had been identified the required number of times. At this time all pairs were again presented with the instructions that S should try as best he could to give the response syllables. The number of correctly identified pairs was used a a measure of learning. S was given a nonsense activity (i.e. connecting numbers on a sheet of paper). After a certain period of time (E_2), S was given a new set of syllables consisting of a certain number of pairs (E_3) to be learned in the same fashion as before. After these pairs were given correctly a certain number of times (E_4), the original fifteen pairs were again presented and the number of pairs correctly recalled was recorded.

The factors were, then,

E_1: degree of original learning
E_2: time in minutes between original and interpolated learning
E_3: length of the interpolated list
E_4: degree of interpolated learning

The criterion, Y, was defined as the percent recall at the end of the experiment relative to the number of syllables given correctly at the end of the original learning.

The First Stage

The factors were coded into standard design factors by:

$$X_1 = (E_1 - 3)/1$$
$$X_2 = (E_2 - 15)/5$$
$$X_3 = (E_3 - 6)/3$$
$$X_4 = (E_4 - 3)/1$$

It was decided to perform a one-half replicate of the 2^4 factorial design with two added center points. At each point of the design, two replications were performed. The center points were added in order to provide for orthogonal blocking in case it was decided to complete the factorial and add axial points for a central composite rotatable design.

The data for this stage are given in Table I. Calculation of (3) gives the fitted equation as

$$Y = 71.75 - 8.56\,X_1 - 9.69\,X_2 - 23.4\,X_3 + 4.94\,X_4$$

The sums of squares due to regression is given by $b'X'Xb$, where b is the vector of estimated partial regression coefficients. Since $X'X$ is a diagonal matrix for this design, each term of the sum is the sums of squares attributable to each regression coefficient. The error sums of squares is found by pooling "within cells sums of squares" where a cell is considered to be a particular treatment combination. The lack of fit sums of squares is found by subtraction. The complete analysis of variance is given in Table II.

Inspection of Table II would suggest X_3 to be influential in determining the amount of retroactive inhibition induced and that the effect of the other factors is negligible. Examination of the data (not shown) revealed that for the treatment combinations where X_1 was -1 (or equivalently, E_1 was 2) the Ss were not learning the original list to the degree we desired. We were interested in the region where the range of learning would vary from two-thirds of the list to an overlearning situation. Further experimentation was undertaken, therefore, in the region where E_1 would be larger.

The Second Stage

The first factor, X_1, was re-coded to

$$X_1 = (E - 5)/1$$

It was decided to perform a half-replicate of the 2^4 factorial using for the trials where $X_1 = -1$ the data already found for $X_1 = +1$ in the first stage. Two center points were added and a double replication of the new design was performed. The results are shown in Table III. The data for the combinations where $X_1 = -1$ is the data for subjects 1-8 in Table I.

The fitted equation was

$$Y = 76.25 + 7.31\,X_1 + 3.19\,X_2 - 17.13\,X_3 - 1.44\,X_4.$$

The analysis of variance calculations proceed as before and are shown in Table IV.

Again we find the dominant factor to be X_3. It was also observed that a satisfactory amount of original learning had occurred. This is reflected in the change of sign of b_1 (which, of course, may be a chance deviation.)

The Third Stage

It was decided to re-code factors E_1 and E_2 using a larger spread of points for these factors to see if a significant effect could be obtained.

Since E_4 was negligible in its effect at both stages, it was decided to drop this factor from further investigation. This factor was set equal to the constant, four, for the balance of the experiment. The equations were

$$X_1 = (E - 6)/2$$
$$X_2 = (E - 15)/15$$
$$X_3 = (E - 6)/3$$

A "cuboctahedron" design with four center points was used. This design, first given by DeBaun (7) and later by Box and Behnken (2), is a "near-rotatable, three-level, second order design. ("Near-rotatability" refers to a design which approximates a rotatable design more closely than the classical factorials). With four center points, orthogonality of the quadratic effects is obtained. The design is shown in figure 5.

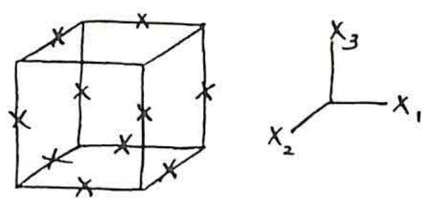


Figure 5: Cuboctahedron Design

Two replications were again performed. The results are given in Table V.

A second-order model was fitted to the data for this stage and also to the data for the second and third stages combined. Typically, one would utilize only the last set of data generated since polynomial approximations over regions of some distance apart are poor. This was not the case here since the cuboctahedron was performed in the same region as the second stage of experiments.

The analysis of variance is given in Table VI.

The sums of squares for a particular coefficient is that portion of the total regression sums of squares attributable to the fitted coefficient after fitting the coefficients preceeding it in Table VI. The sum of squares for the q^{th} coefficient adjusted for all the others can be calculated from b_q^2/C_{qq}, where b_q is the estimated partial regression coefficient and C_{qq} is the term in the q^{th} row and q^{th} column of the inverse, $(X'X)^{-1}$. The contribution of a variable, X_2 for example, adjusted for all the others can be calculated from $b^{*\prime} C^{*-1} b^*$, where

$$b^{*\prime} = (b_2, b_{22}, b_{23})$$

and C^{*-1} is the inverse of the matrix formed by selecting the rows and columns of $(X'X)$ which correspond to the regression coefficients in b^*. The F-values for the variable, X_2, and the coefficients, b_{33} and b_{13}, were non-significant and consequently were dropped from the model.

The final equation was -

$$Y = 86.58 + 6.21 X_1 - 15.59 X_3 - 9.12 X_1^2.$$

Transforming from the design variables back to the original variables, E_1 and E_3, gives

$$E_1 = 6.68 \text{ and } E_3 = 22.86 - Y/5.20.$$

The contours of response, then, are parabolae with foci on the line $E_1 = 6.68$. The response surface is shown in Figure 6. Within the region of experimentation, there appears to be a maximum with respect to E_1, but with respect to E_3 the surface is a "rising ridge". This is interpreted to signify that recall is maximized at a certain intermediate level of degree of learning and with successively shorter and shorter lengths of the interpolated list. Note that for Y = 100% recall, E_3 is near zero. Time, E_2, was, of course, not expected to be significant.

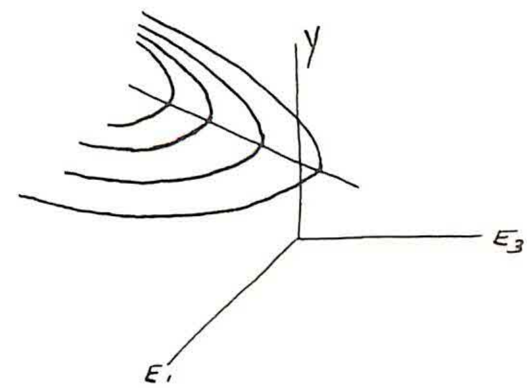


Figure 6: Response Surface

Summary

An overview of response surface methodology has been presented together with an application to a human learning experiment. Readers desiring more detailed discussions are referred especially to references 1, 4, 5, 6. The major portion of the mathematical and statistical theory is given in reference 3.

REFERENCES

1. Box, G. E. P., "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples," Biometrics, X (1954), pp. 16-60.
2. Box, G. E. P. and Behnken, D. W., "Some New Three Level Designs for the Study of Quantitative Variables," Technometrics, II (1960), pp. 455-76.
3. Box, G. E. P. and Hunter, J. S., "Multifactor Experimental Designs for Exploring

TABLE VI

ANALYSIS OF VARIANCE FOR THIRD STAGE

Source	d.f.	SS	MS	F
Due to regression	10	352389.567		
b_0	1	341172.000		
b_1	1	1357.592		6.17
b_2	1	342.224		1.56
b_3	1	7781.281		35.37
b_{11}	1	968.575		4.40
b_{22}	1	17.750		.08
b_{33}	1	77.684		.35
b_{12}	1	268.926		1.22
b_{13}	1	313.504		1.43
b_{23}	1	90.031		.41
About regression	42	9830.433		
Lack of fit	13	3451.433	265.495	
Error	29	6379.000	219.495	1.21
Total	52	362220.000		

Response Surfaces, "Annals of Mathematical Statistics, XXVIII (1957), pp. 195-241.

4. Box, G. E. P. and Youle, P. V., "The Exploration and Exploitation of Response Surfaces: An Example of the Link Between the Fitted Surface and The Basic Mechanism of the System," Biometrics XI (1955), pp. 287-323.
5. Chew, V., editor, Experimental Design in Industry (1958) New York: John Wiley and Sons.
6. Cochran, W. G. and Cox, G. M., Experimental Designs, Second Edition (1957) New York: and Sons.
7. DeBaun, Robert M., "Response Surface Designs for Three Factors at Three Levels," Technometrics I (1959) pp. 1-8.
8. Glaze, J. A., "The Association Value of Nonsense Syllables," Journal of Genetic Psychology, XXXV (1928) pp. 225-267.
9. McGeoch, J. A. and Irion, A. L., The Psychology of Human Learning, (1952) New York: Longmans, Green and Co.
10. Meyer, Donald L., Response Surface Designs with Integer-Valued Factors. (1961) Unpublished PhD dissertation, University of Minnesota.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

THE USE OF A CROSS-OVER DESIGN IN A STUDY OF STUDENT TEACHERS' CLASSROOM BEHAVIORS[1]

ROGER E. WILK[2]
University of Minnesota

ARE APPRAISALS OF students made at the time of their admission to the teacher education program related to descriptions of their subsequent classroom behavior? This question stimulated a study of the observed classroom behaviors of women in elementary education during their student teaching (9). This paper describes that part of the study which was concerned with the development of the experimental design for studying classroom behaviors. Since few, if any, studies of student teachers' classroom performance have employed an experimental approach, this discussion is presented as an example of the application of the experimental method to this type of research and of the use of the cross-over design in particular. The study illustrates how factors assumed to be relevant to classroom performance variables can either be controlled or their influence measured and examined for significance.

Sources of Variability Considered

To be able to answer the general question which introduces this paper, it was necessary to identify and arrange the factors relevant to the observations of elementary student teachers in such a way that the impact of the personality of the individual student teacher could be separated from situational and experiential influences. The design needed to control some factors, preventing unknown biases, and to estimate the influences of other factors that could be varied experimentally. A study of the practices regarding the assignment of elementary student teachers in the College of Education led to the identification of the sources of variability that were considered. They are phrased in the following question:

What is the relationship of preference for student teaching at a particular grade level, student teaching experience, and the classroom situation to the observed behavior of student teachers?

The investigators selected the factors of grade level preference and student teaching experience for study since they were relevant to the operation of the student teaching program and were, presumably, related to classroom behaviors to be observed. At the time of the study, student teaching placements were made after considering the preference that a student had for doing student teaching at a particular grade level. Though students could not always be assigned to the grade level that was the first preference, it was assumed that they would be more successful if the initial quarter could be in a grade near the preferred level. The typical student teaching experience for students in the Elementary Education IA curriculum was two quarters of half-day student teaching at two different grade levels. The grade level in which the student teacher was placed and the quarter of student teaching, whether the first or second quarter of student teaching, would undoubtedly affect the observed behaviors. In addition, the order or sequence in which the two grade levels were assigned for the two quarters might also affect the observed behaviors. Thus the three related student teaching experience factors, grade level, quarter, and order of experience, plus the factor of grade level preference became the four experimental variables for the study.

The effects of the classroom situational variables were not studied, but they were controlled. The accessibility of the school from the student teacher's residence, and the socio-economic level of the school neighborhood were controlled by dividing the Twin Cities into four geographical areas. The boundaries of these areas were tentatively set so that the schools would be accessible to student teachers living in the area. To control the socio-economic conditions, the boundaries of these areas were adjusted so that the characteristics of the school neighborhoods, measured by a socio-economic index derived from census data, were comparable from area to area. Further control was exercised through the selection of the sample. All student teachers were women in the Elementary Education IA curriculum, and in addition to having completed two of her three college years in

the College of Education at the University of Minnesota, each one had also participated in a previous study (3). All of the subjects were given morning assignments during the fall and winter quarters.

Figure 1 illustrates the arrangement of the four experimental factors in the cross-over or switch-over design. Subjects eligible for placement in the experiment were first stratified into two groups according to their stated grade level preference, those who preferred to teach upper grades (4, 5, 6) designated prefer U, and those who preferred to teach lower grades (1, 2, 3) designated prefer L.

Preference	Quarter 1	Quarter 2
Prefer U	U	L
	L	U
Prefer L	U	L
	L	U

Figure 1 Cross-Over Design

Each of these preference groups was randomly assigned to one of the two orders or sequences of experience. One-half of each preference group received its stated choice the first quarter; the other did not. The model specifying any individual observation can be stated thus:

$X(ijk\ell)$ is the observation score for the ℓth student teacher with the ith preference, teaching at the jth level, assigned the kth order so that

- $i = 1$: prefer upper grades (4, 5, 6)
- $i = 2$: prefer lower grades (1, 2, 3)
- $j = 1$: upper grade level (4, 5, 6)
- $j = 2$: lower grade level (1, 2, 3)
- $k = 1$: upper-lower order
- $k = 2$: lower-upper order
- $\ell = 1, 2, 3, \ldots$ identifies the student teacher.

It is assumed that $X(ijk\ell)$ are random variables, normally distributed with the following specifications:

$E[X(ijk\ell)] = \omega(ijk)$
Variance $[X(ijk\ell)] = \sigma^2$
Covariance $[X(ijk\ell), X(i'j'k'\ell')] = \rho\sigma^2$
for $i = i'$, $j = j'$, $k = k'$, $\ell = \ell'$, otherwise zero.
$\omega(ijk)$ is further defined as a linear function of fixed parameters which are representative of the various factors affecting a score. Thus

$$\omega(ijk) = \mu + P(i) + L(j) + O(k) + PL(ij) + LO(jk) + PLO(ijk)$$

The analysis of variance for this cross-over design is presented in Table I. This analysis follows the pattern given by Cochran and Cox (2) and Stanley (8) for counterbalanced observations. The sources of variation are separated into two sets, those that are concerned with "within individuals" differences (Set 1) and those concerned with "between individuals" differences (Set 2), and each set designates an error term appropriate for testing hypotheses within a set. As is usual in this type of design and the present analysis, the assumption of no treatment carry-over for a subject is made.

Classroom Behaviors Observed

A number of classroom behavior variables were measured, but only two are presented here for illustrative purposes, need Dominance (n Dom) and need Integration (n Intg). These concepts, drawn primarily from the work of Murray (7) and H. H. Anderson (1), were basic to this study and the previous investigation upon which this study was built (3). In general the definitions of these variables state that the teacher who has a high degree of n Dom is one who directs learners toward specific, designated goals through a closely regulated set of procedures. A teacher having a high degree of n Intg is one who supports learners toward a wider variety of goals using a flexible set of procedures.

To assess n Dom and n Intg trained observers time-sampled and categorized the student teachers' verbal behaviors according to the Minnesota System of Interaction Analysis developed by Flanders (4). Verbal behaviors categorized by this system as direct influence statements, those which restricted pupils' freedom to respond, and indirect influence statements, those that increased the freedom of the pupils to respond, were accepted as behavioral evidence of n Dom and n Intg respectively. Each of five observers, for whom high inter-observer agreement was established prior to the experiment, visited each student teacher each quarter and categorized three five-minute samples of her verbal behavior. A student teacher's n Dom (RA) score for one quarter was the composite of the five different observer's scores, each one of which was the weighted sum of those verbal statements categorized as direct influence attempts. The weighting of scores from each visit was accomplished by the use of the reciprocal averages (RA) technique (6), (5). Similarly, the n Intg (RA) scores were the weighted sums of the indirect influence attempts recorded on one observation summed over the five observations. The results of the analysis of the n Dom (RA) and n Intg (RA) scores using the cross-over design follow.

Results and Conclusions

Data were analyzed from thirty-four women who were majors in the Elementary IA curriculum in the College of Education at the University of Minnesota who were the subjects of the study. Each

TABLE I

ANALYSIS OF VARIANCE FOR COUNTERBALANCED OBSERVATIONS OF STUDENT TEACHERS AT BOTH UPPER AND LOWER ELEMENTARY GRADES OVER A TWO QUARTER PERIOD STRATIFIED ACCORDING TO GRADE LEVEL PREFERENCE

SV	df
Set 1	
L (j) (level)	L - 1
LO (jk) (Quarter)	(L - 1)(O - 1)
PL (ij)	(P - 1)(L - 1)
PLO (ijk)	(P - 1)(L - 1)(O - 1)
Error (Set 1)	PO(L - 1)(K - 1)
Total (Set 1)	POK(L - 1)
Set 2	
P (i) (preference)	P - 1
O (k) (order)	O - 1
PO (ik)	(P - 1)(O - 1)
Error (Set 2)	PO(K - 1)
Total (Set 2)	POK - 1

TABLE II

ANALYSIS OF VARIANCE OF NEED DOMINANCE (RA) SCORES FOR COUNTERBALANCED OBSERVATIONS OF STUDENT TEACHERS AT BOTH UPPER AND LOWER ELEMENTARY GRADES OVER A TWO QUARTER PERIOD STRATIFIED ACCORDING TO GRADE LEVEL PREFERENCE

SV	df	SS	MS	F
Set 1				
L (j) (levels)	1	986.4853	986.4853	11.2750**
LO (jk)	1	718.2498	718.2498	8.2092**
PL (ij)	1	19.5223	19.5223	<1
PLO (ijk)	1	83.4544	83.4544	<1
Error (Set 1)	30	2624.7882	87.4929	
Total (Set 1)	34	4432.5000		
Set 2				
P (i) (preferences)	1	7.5436	7.5436	<1
O (k) (order)	1	440.1324	440.1324	4.5609*
PO (ik)	1	.0625	.0625	<1
Error (Set 2)	30	2895.0304	96.5010	
Total (Set 2)	33	3342.7689		

*Significant at .05 level.
**Significant at .01 level.

student teacher was observed by all five observers during each of two quarters in the classroom where she had been assigned according to the cross-over design.

Table II presents the tests of significance for the n Dom (RA) scores. These tests indicate that three F values were significant. All three refer to main effects if level by order is considered a quarter difference. Level and level by order are within individual differences, but order differences occur between individuals. The mean Dom (RA) scores for these tests are as follows:

order	level		
	upper	lower	both levels
upper-lower	40.5	41.6	41.1
lower-upper	39.1	53.2	46.2
both orders	39.8	47.4	

The table of means indicates that n Dom (RA) scores were significantly higher for the group of students whose order first placed them in the lower grades then moved them to the upper grades for the second quarter of experience. It also indicates that for all students the n Dom (RA) scores were higher in the lower grades than in the upper grades. An inspection of the four level by order means suggests that the differences among the four cannot be interpreted as a between quarters difference. The mean score (53.2) of the group who had the upper-lower order when they were at the lower level (first quarter) is singly responsible not only for the significant interaction, but for the two main effects as well.

Table III indicates that in the analysis of n Intg (RA) scores the only significant difference was found for the interaction of level by order (quarter). In this design the level by order and the quarter effects were confounded. Reference to the set of n Intg (RA) means below indicates that differences do exist but cannot be interpreted either as a quarter effect or a level by order effect.

order	level	
	upper	lower
upper-lower	33	58
lower-upper	56	38

Although the primary purpose of this paper is illustrative, the analyses of data do warrant one important conclusions. The experimental factors related to the student teaching experience did influence the student teachers' verbal behaviors. Student teachers who taught first in the lower grades were more direct in their influence attempts when they taught in the lower grades than when in the upper grades, but this effect did not occur for those who began student teaching in the upper grades. In addition, student teachers showed more indirect influence attempts during their second quarter than the first regardless of other factors. There was no evidence that grade level preference influenced either n Dom or n Intg (RA) scores. These results will be combined with the results of the analyses of other observation variables and treated more fully in subsequent papers.

Summary

The purpose of this paper is to report a study which illustrates a use of the experimental method in studying the classroom behaviors of teachers. The cross-over design was employed to place student teachers in classrooms in such a way that the influence of their grade level preference and their student teaching experience could be estimated. The paper also describes the way in which the school location and the socio-economic level of the school neighborhood was treated to prevent these factors from introducing an unknown bias in this results.

Thirty-four women who were Elementary IA majors in the College of Education at the University of Minnesota and a part of a prior study were visited by five different observers during each of their two quarters of student teaching. Using the Minnesota System of Interaction Analysis, the observers time-sampled and classified teachers' verbal behaviors according to the System. The direct and indirect influence attempts were weighted and totaled separately to give measures of need Dominance and need Integration, respectively. Significant differences found in the analysis of n Dom (RA) scores were attributed to the fact that student teachers who taught the first quarter in the lower grades and the second quarter in the upper grades had significantly higher scores when they were teaching in the lower grades. Analysis of the n Intg (RA) scores showed that student teachers had higher scores the second quarter than they did the first. It was concluded that grade level taught, experience in student teaching, and the order of the student teaching experience did affect the observed behaviors of student teachers. A determination of the influence of these factors would not have been possible without the use of an experimental design which made it possible to assign a causal effect to significant factors.

FOOTNOTES

1. The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.
2. The helpful assistance of Raymond O. Collier, Jr. and Donald G. MacEachern in the development of the design of the study is gratefully acknowledged.

REFERENCES

1. Anderson, H. H. "The Measurement of Domination and of Socially Integrative Behavior in Teachers' Contacts with Children." Child Development, (1939) X, pp. 73-89.
2. Cochran, W. S. and Cox, Gertrude M. Experimental Designs, (1950) New York: John Wiley and Sons.
3. Edson, W. H. and Wilk, R. E. An Experimental Study of the Admissions Interview in Teacher Education for Predicting Success in Teaching. (Multilith) College of Education, (1958) University of Minnesota.
4. Flanders, N. A. "Teacher Influence, Pupil Attitudes and Achievement," (Multilith) College of Education (1960) University of Minnesota.
5. Mitzel, H. E. and Hoyt, C. J. "A Methodological Study of Reciprocal Averages Study Technique Applied to an Attitude Scale," Journal of Counseling Psychology (1954) I, pp. 256-59.
6. Mosier, C. J. "Measurement in Rural Housing, A Preliminary Report," Educational Psychological Measurement (1942), pp. 139-52.
7. Murray, H. A. Explorations in Personality, (1938) New York: Oxford University Press.
8. Stanley, J. C. "Statistical Analysis of Scores from Counterbalanced Tests," The Journal of Experimental Education (1955) XXIII, pp. 187-207.
9. Wilk, R. E. and Edson, W. H. A Study of the Relationships Between Observed Classroom Behaviors of Elementary Student Teachers, Predictors of Those Behaviors, and Ratings by Supervisors. (Multilith) College of Education (1962) University of Minnesota.

TABLE III

ANALYSIS OF VARIANCE OF NEED INTEGRATIONS (RA) SCORES FOR COUNTERBALANCED OBSERVATIONS OF STUDENT TEACHERS AT BOTH UPPER AND LOWER ELEMENTARY GRADES OVER A TWO QUARTER PERIOD STRATIFIED ACCORDING TO GRADE LEVEL PREFERENCE

SV	df	SS	MS	F
Set 1				
L (j) (levels)	1	194.4853	194.4853	2.087
LO (jk)	1	7687.1911	7687.1911	82.4864**
PL (ij)	1	13.7039	13.7039	<1
PLO (ijk)	1	143.3166	143.3166	<1
Error (Set 1)	30	2795.8031	93.1934	
Total (Set 1)	34	10834.5000		
Set 2				
P (i) (preferences)	1	27.9225	27.9225	<1
O (k) (order)	1	51.1912	51.1912	<1
PO (ik)	1	240.9980	240.9980	1.712
Error (Set 2)	30	4223.1972	140.7732	
Total (Set 2)	33	4543.3089		

* Significant at .05 level.
** Significant at .01 level.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

EXPOSITORY NOTES ON THE PROBLEM OF MAKING MULTIPLE COMPARISONS IN A COMPLETELY RANDOMIZED DESIGN

JACK N. SPARKS
Colorado State College

MANY OF the current tests and reference books used by educational researchers are quite sharp in their condemnation of the practice of performing t-tests on each pair of means as a method of testing the hypothesis that, in a completely randomized design, $\mu_1 = \mu_2 = \ldots = \mu_1$ when $k > 2$. The practice generally recommended as a suitable alternative, however, has but little more merit.

The authors of these works quite rightly indicate that when a given critical value is set for testing all the possible pairs of differences between a number of means in this way, the probability of at least one false significant difference will be greater than the critical value. The practice sometimes suggested and followed today involves applying the F-test for analysis of variance and following that with t-tests for all possible comparisons between means. Such a procedure ignores one of the principal arguments against applying the t-test in the first place. As Steel and Torrie (19) point out, with three treatments the observed value of "t" for the greatest difference will exceed the tabled .05 level about 13% of the time even if the treatment populations truly do not have differing means. This change in significance level is progressive, and for six treatments the value of "t" for the greatest difference will be greater than the tables .05 level about 40% of the time.

It is true that some protection against the error of rejecting a true hypothesis of no difference is provided if specific comparisons are made only after the application of the F-test and only if the result indicates rejection of the hypothesis that $\mu_1 = \mu_2 = \ldots \mu k$. It would seem, however, that at this point it would be appropriate to turn to one of a number of procedures for specific comparisons among treatment means which have been developed in recent years. These procedures, in general, are designed to set an experiment-wise error rate or level of significance.

A number of the more apparently useful methods of comparison have received some attention in recent years (5, 9, 10, 11, 15, 19, 22).

In an article in the September 1955 issue of Psychological Bulletin, McHugh (9) discussed the problem of examining comparisons between means which are suggested by the data. A method proposed by Scheffé (13) was applied to two examples of this type of situation. Additional interest in such a posteriori or "post-mortem" analyses was stirred by a paper prepared by Stanley (15) and also published in the Psychological Bulletin. Stanley pointed out that Scheffé's test, while offering an experiment-wise level of significance, was not as powerful as some other tests proposed. A procedure developed by Tukey (20) was offered as an alternative. For the instance in which comparisons of each experimental group with a control only are desired, Stanley offered the procedure suggested by Dunnett (4).

More recently, Thomas A. Ryan (11) discussed the questions of logic which arise in choosing methods of dealing with multiple comparisons and other multiple statistical tests. He contended that the decision with respect to which methods of comparison are correct should not revolve around whether or not the comparisons being made were specified in advance of data collection (a priori) or were suggested by the data (a posteriori). Rather, wrote Ryan, the fact is that the classical methods are inappropriate for multiple comparisons under any circumstances, and should be replaced by the newer methods in either instance. It should be noted that Ryan also presented a very comprehensive discussion of the considerations involved in determining the actual error rate or level of significance in a given experiment involving multiple comparisons. A resulting exchange by Gaito (6) and Ryan (12) resulted in further clarification of the problem and proposed solutions. Nevertheless, the question of whether the appropriate error rate should be per comparison or per experiment continues to be debated (13, 21), with the weight of evidence seeming to fall on the side of those who espouse the use of an experiment-wise error rate.

One point raised by Ryan (12) with respect to the newer methods for multiple comparisons should be made quite clear. The Scheffé, Tukey, Dunnett and certain other of the available methods may be applied without the use of an initial F-test of the

overall difference between groups. It should be added, however, that it will often be desirable to apply such a preliminary test to determine whether further analysis is even worthwhile.

Description of Procedures

In this section, four of the newer procedures will be described in terms of their application to the completely randomized design. Some of these procedures were developed for making differences between pairs of means only, while others are appropriate for the purpose of testing all-linear contrasts. However, since the interest of most researchers in education is in the differences between pairs of mean only, the presentation is specific to tests about these particular linear contrasts.

A. Duncan's Multiple Range Test

Duncan (2) originally developed this test for the situation in which there were equal numbers of subjects for all treatments. Under this condition the steps in the procedure are:

1. Calculate an estimated standard error of the mean ($\tilde{\sigma}_{\overline{X}}$) based upon the mean square within treatments (ms_W):

$$\tilde{\sigma}_{\overline{X}} = \sqrt{\frac{ms_W}{n}}$$

where "n" is the number of subjects in each treatment. This standard error will have degrees of freedom equal to that of ms_W.

2. Locate significant studentized ranges for the .05 or .01 level given in the table provided by Duncan (2). The table is entered at the row for the degrees of freedom for ms_W and significant studentized ranges are found for $p = 2$ to k, the sizes of "p" indicating the number of means involved in a comparison. Values not in the table can be found by interpolation.
3. Multiply the significant ranges obtained above by $\tilde{\sigma}_{\overline{X}}$ to give the least significant ranges.
4. Rank the means.
5. Compare first the difference between the largest and smallest means with the least significant range calculated for the largest number of means in a comparison (since this difference involves comparison over all means). Continue to test the differences in the order indicated:

   Largest -- second smallest
   Largest -- third smallest
   . .
   . .
   . .
   Largest -- second largest
   Second largest -- smallest
   Second largest -- second smallest
   . .
   . .
   . .
   Second smallest -- smallest

   Thus, all possible comparisons have been made.
6. The differences which exceed the corresponding least significant range are declared significant. The only exception to this is that no two means can be said to be significantly different if they are contained in a larger subset of means which has a non-significant range.

Kramer (7) has proposed a method for extension of Duncan's test to the situation in which there are unequal numbers of subjects under the various treatments. The procedure dictates the following changes:

1. Obtain the significant studentized ranges from the table and multiply by

$$\sqrt{ms_W} \quad \text{rather than} \quad \sqrt{\frac{ms_W}{n}}$$

2. For specific comparisons, multiply the appropriate intermediate value as obtained above by

$$\sqrt{1/2 \quad \frac{1}{n} + \frac{1}{n_j}}$$

where "i" and "j" represent the treatments being compared.

The validity of Kramer's extension has not been adequately established however. As Cochran and Cox (1) point out, Duncan's multiple range test makes the level of protection against a Type I error .95 for sets of two means when the level of significance is .05. For sets of three means, however, the level is decreased to $(0.95)^2$ or 9027; for sets of four means the level is only .857. This decrease in the level of protection makes the multiple range test rather sensitive to detecting real differences, but still at the expense of lowered protection against the error of rejecting a true hypothesis of no difference.

B. Tukey's W-Procedure (20)

This procedure was also originally developed for the situation in which there are equal numbers of subjects under all treatments. It gives a single value for judging the significance of all differences, and the significance level holds for the experiment as a whole. This means that when the .05 level is set, 5% of such experiments will result in one or more false significant differences. The procedure is as follows:

1. Compute $w = {}_{\alpha}q_{(p,\,v)}\,\tilde{\sigma}_{\overline{X}}$, where "q" may be obtained from the table for the upper percentage points of the studentized range prepared by May,

(8) "p" is the number of treatments, "α" is the level of significance, "v" is the number of degrees of freedom for mean square within treatments (ms_W), and $\tilde{\sigma}_{\overline{X}}$ is

$$\sqrt{\frac{ms_W}{n}}$$

2. Use "w" to judge the significance of each of the observed differences between means.

An extension of Tukey's procedure to the case of unequal numbers of subjects under the various treatments has been proposed. The following changes are necessary (19):

1. Calculate $w' = {}_{\alpha}q\ (p, n)\ \sqrt{ms_W}$
2. To make any desired comparison between means, multiply w' by

$$\sqrt{1/2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Usually, fewer differences will be found to be significant by Tukey's procedure than by Duncan's method. The reason is made obvious when we recall that Tukey's procedure provides that in 95% of our experiments we will declare no differences significant if there are actually no differences between treatment population means and that a single value is used for determining the significance of all observed differences between means.

No adequate discussion of the effect of non-normality and heterogeneity of treatment groups variances upon these multiple range tests is available. Nevertheless, they are preferable to the use of t-tests of all possible differences. Nonparametric analogues of the Duncan and the Tukey procedures are discussed in a recent article by Steel (18).

C. Scheffé's S-Method

Scheffé (14) has developed an "S-test" for multiple comparisons which he claims to be insensitive to violations of the conditions of normality and equality of variance. It is appropriate not only for comparisons involving differences between pairs of means, as is our interest here, but also for more complicated contrasts between and among the means. The procedure for comparisons between pairs of means is as follows:

1. Compute

$$S^2 = \frac{(\overline{X}_i - \overline{X}_i)^2}{ms_W\ \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

where $\overline{X}_i$ and $\overline{X}_j$ are the means of the treatment groups being compared, and "n" is the number of the subjects in the subscripted treatment group.

2. Each S^2 should now be compared with

$$(k - 1)\ {}_{\alpha}F_{k-1, N-k}$$

Each S^2 larger than this standard indicates a significant difference between means.

3. Where all subgroups are of the same size, n_i always equals n_j and a critical difference ($d = \overline{X}_i - \overline{X}_j$) can be computed for any given level of significance (α). Thus,

$$d = {}_{\alpha}S_{k-1,\ N-k}\ \sqrt{\frac{2\ ms_W}{n}}$$

The procedure outlined above is an adaptation which is suitable only for simple comparisons of differences between means. For more involved contrasts the reader is referred to Scheffé's original article.

The Scheffé test is so designed that all possible linear combinations of means may be investigated, while the selected level of significance is retained. This is an advantage if one is interested in all of those possibilities, but it results in lowered power if the interest is in only some of these combinations. In this paper, however, our interest is in only certain linear contrasts—the differences between all possible pairs of means. It would seem that under these conditions, the Tukey test, which was originally developed to apply only to differences between pairs of means, would be somewhat more powerful than Scheffé's test. As a matter of fact, Scheffé himself indicates that this is true (15).

This does not mean, however, that the Tukey test is always the one to use in our situation. Olive Dunn (3) had developed a procedure for testing hypotheses about linear contrasts among means which has more power than either the Tukey or Scheffé procedure, provided the experimenter is willing to specify in advance a limited number of combinations. This procedure also has the advantage over Tukey's procedure of not being limited by conditions with respect to homogeneity of variance.

D. Dunn's "c" Procedure

As indicated, Dunn's procedure is more powerful than the Scheffé S-test only if the experimenter is willing to restrict the number of linear comparisons (m) among k means. The procedure is based upon the student t-distribution, and the power of the test depends upon m, the actual number of linear contrasts being examined, while the power of the Scheffé test depends upon k, the number of means being compared. The Dunn procedure is more powerful for small m and large k, and the Scheffé procedure is more powerful for large m and small k. Dunn's article describing the procedure includes a number of tables demonstrating how it is possible in any given case to determine beforehand which procedure will be more powerful.

The steps in applying the procedure in testing hypotheses concerning differences between means (as adapted from Dunn's description) are as follows:

1. Compute

$$c = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{ms_w\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

for each pair of means to be computed. $\overline{X}_i$, $\overline{X}_j$, ms_w, n_i, and n_j are as previously defined.

2. Dunn's article contains tables (Table I and Table II) of c for σ = .05 and .01 with selected values of m, the total number of comparisons being made, and v, the degrees of freedom of ms_w. Each of the values of c obtained in (1) above should be compared with the tables c. Each c larger than this tabled value indicates a significant difference between means.
3. For the case in which all subgroups are of the same size, a critical difference (d = $\overline{X}_i - \overline{X}_j$) may be computed for any given level of significance, α. Thus

$$d = {}_{\alpha}c_{m,v}\sqrt{\frac{2\,ms_w}{n}}$$

It should be apparent now, that this paper will not designate a specific one of these procedures as the right one for adoption. The research worker is obligated to decide, on the bases of information which is available to him before the experiment is performed, which procedure will best fit his needs. Dunn (3) and Scheffé (14) both present some interesting tables relative to comparisons of this nature.

An Example Using Hypothetical Data

Four methods of teaching percentage to sixth graders (case method, formula method, equation method, and unitary analysis method) were devised, and detailed lesson plans were prepared for each. Twenty-eight sixth-grade classes were available for an experiment, one class in each of twenty-eight elementary schools. The classes were randomly assigned to four groups, and the four groups were randomly assigned to the four treatment methods. Thus, seven classes were taught percentage by each method. Fourteen consecutive school days were allotted for instruction and teachers were provided with the detailed lesson plans for each day. A criterion test which had been tested in a pilot project was prepared and administered at the end of the experiment. Table I presents the mean for each class on the criterion test. Since the class was the sampling unit, the class must be the unit of analysis, and the mean class score will be the criterion measure.

The first test involved the test of the overall hypothesis that the means of the hypothetical treatment populations are equal ($\mu_1 = \mu_2 = \mu_3 = \mu_4$). An analysis of variance yielded the results presented in Table 2.

The obtained F was compared with the tabled F value for 3 and 24 degrees of freedom to determine whether or not the null hypothesis should be rejected. Since the tabled F value at the .05 level is only 3.01, the hypothesis was rejected.

Since it is apparent at this point that further analyses are in order to determine which treatment populations can be said to have different means, each of the three methods described in this article will be used for this purpose.

A. Duncan's Multiple Range Test

1. $$\tilde{\sigma}_{\overline{X}} = \sqrt{\frac{ms_w}{n}} = \sqrt{\frac{17.37}{7}} = 1.57$$

2. Significant studentized ranges for the number of means to be compared;

number of means (p)	2	3	4
significant studentized range	2.92	3.07	3.15

3. Least significant ranges ($\tilde{\sigma}_{\overline{X}}$ times significant studentized ranges)

number of means	2	3	4
least significant range	4.58	4.82	4.95

4. Ranked means:

a. Unitary analysis method	32.52
b. Equation method	27.84
c. Formula method	21.25
d. Case method	18.68

5. Table of differences between means:

	d	c	b
a.	13.48*	11.27*	4.68*
b.	9.16*	6.59*	
c.	2.57		

6. The differences which exceed the corresponding least significant ranges are marked with an asterisk. The mean for the group taught under the unitary analysis method is significantly larger than all others. In addition, the equation method would be said to be better for this population than the formula or case method, but the latter two methods have not been demonstrated to produce different results.

B. Tukey's W-Procedure

1. For the .05 level of significance $q_{4,24} = 3.90$

TABLE 1

MEAN SCORE OF EACH CLASS ON THE CRITERION TEST

	Case Method	*Formula Method*	*Equation Method*	*Unitary Analysis Method*
	14.59	20.27	27.82	33.16
	23.44	26.84	24.92	26.93
	25.43	14.71	28.68	30.43
	18.15	22.34	23.32	36.43
	20.82	19.49	32.85	37.04
	14.06	24.92	33.90	29.76
	14.26	20.20	23.42	33.88
Mean	18.68	21.25	27.84	32.52

TABLE 2

ANALYSIS OF VARIANCE FOR DATA IN TABLE 1

Source of Variation	df	Sum of Squares	Mean Square	F
Among Methods	3	830.19	276.73	15.93
Within Method	24	416.76	17.37	
Total	27	1246.95		

$w = 3.90 \times 1.57$
$= 6.12$

2. The table of differences between means as previously set forth is compared with "w"

	d	c	b
a.	13.84*	11.27*	4.68
b.	9.16*	6.59*	
c.	2.57		

3. Using this test one less difference is declared significant than under the Duncan Test. The unitary analysis and equation methods would not be said to differ in effectiveness for this population.

C. Scheffé's S-Test

1. The ${}_{.05}F_{3,\ 24}$ is 3.01. This value must be multiplied by (k-1) = 3 to obtain the critical S^2.
2. Since all subgroups are of equal size, a critical difference procedure may be used:

$$d = {}_{.05}S_{3,24} \sqrt{\frac{2\ ms_w}{n}}$$

$$= \sqrt{9.03} \quad \sqrt{4.99}$$

$$= 6.71$$

3. The table of differences between means is as before:

	d	c	b
a.	13.84*	11.27*	4.68
b.	9.16*	6.59	
c.	2.57		

4. As we might have predicted, Scheffé's test produced fewer significant differences than did Tukey's. The best we can say here is that the unitary analysis method and the equation method are better for this population than the case method, and the unitary analysis method also seems to be better than the formula method.

D. Dunn's c-procedure

1. The tabled value of c for $\alpha = .05$ (Table 1 in Dunn, reference 3) and m=6, v= 24 is 2.88.
2. Since all subgroups are of the same size, a critical difference may be calculated:

$$d = 2.88 \sqrt{\frac{2\ ms_w}{n}}$$

$$= 6.42$$

3. The table of differences between means:

	d	c	b
a.	13.84*	11.27*	4.68
b.	9.16*	6.59*	
c.	2.57		

4. We would here declare the unitary analysis and equation methods are likely to be superior to the other two methods for this population.

The Scheffé test and the Dunn test have the same experiment-wise error rate, but in this situation the Dunn procedure declares more differences to be significant.

Summary

1. The practice of using t-tests to make multiple comparisons following the significant outcome of an F test in the case of a completely randomized design has been criticized.
2. Proposals for more appropriate analyses have been reviewed.
3. Applications of four suggested procedures to certain linear contrasts in this design have been set forth and examples have been given.

Bibliography

1. Cochran, W. G. and Cox, G. M. Experimental Designs. New York: John Wiley and Sons, Inc., 1957, 76.
2. Duncan, D. B. "Multiple Range and Multiple F-tests." Biometrics, 1955, 11: 1-42
3. Dunn, Olive J. "Multiple Comparisons Among Means." Journal of the American Statistical Association, 1961, 56: 52-64.
4. Dunnett, C. W. "A Multiple Comparison Procedure for Comparing Several Treatments with a Control." Journal of the American Statistical Association, 1955, 50: 1096-1121.
5. Edwards, A. L. Experimental Design in Psychological Research, revised edition. New York: Holt, Rinehart and Winston, 1960.
6. Gaito, J. "Multiple Comparisons in Analysis of Variance." Psychological Bulletin, 1959, 56: 26-47.
7. Kramer, C. Y. "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications." Biometrics, 1956, 12: 307-310.
8. May, J. M. "Extended and Corrected Tables of the Upper Percentage of the 'Studentized' Range." Biometrika, 1952, 39: 192-193.
9. McHugh, R. B. and Ellis, D. S. "The 'Post-Mortem' Testing of Experimental Comparisons." Psychological Bulletin, 1955, 52: 425-28.
10. McNemar, Quinn. Psychological Statistics, third edition. New York: John Wiley and Sons, 1962.
11. Ryan, T. A. "Multiple Comparisons in Psychological Research." Psychological

Bulletin, 1959, 56: 26-47.
12. Ryan, T. A. "Comments on Orthogonal Components." Psychological Bulletin, 1959, 56: 394-96.
13. Ryan, T. A. "The Experiment as the Unit for Computing Ratio of Error." Psychological Bulletin, 1962, 59: 301-304.
14. Scheffé, H. "A Method for Judging All Contrasts in the Analysis of Variance." Biometrika, 1953, 40: 87-104.
15. Scheffé, H. The Analysis of Variance. New York: John Wiley and Sons, Inc., 1959.
16. Stanley, J. C. "Additional 'Post-Mortem' Tests of Experimental Comparisons." Psychological Bulletin, 1957, 53: 128-30.
17. Steel, R. G. D. "Answer to Query 136." Biometrics, 1961, 17: 326-28.
18. Steel, R. G. D. "Some Rank Sum Multiple Comparisons Tests." Biometrics, 1961, 17: 539-552.
19. Steel, R. G. D. and Torrie, J. H. Principles and Procedures of Statistics. New York: McGraw-Hill Book Company, Inc., 1960.
20. Tukey, J. W. "The Problem of Multiple Comparisons." Dittoed, Princeton University, 1953.
21. Wilson, Warner. "A Note on the Inconsistency Inherent in the Necessity to Perform Multiple Comparisons." Psychological Bulletin, 1962, 59: 296-300.
22. Wolf, F. L. Elements of Probability and Statistics. New York: McGraw-Hill Book Company, Inc., 1962.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

A NOTE ON THE MULTIPLE REGRESSION TECHNIQUE FOR DELETING VARIABLES IN THE DISCRIMINANT FUNCTION

RAYMOND O. COLLIER, JR.
University of Minnesota

IN AN EARLY PAPER on taxonomic problems, Fisher, [1, 184] showed that the linear discriminant function for two groups can be conveniently obtained by constructing a certain multiple regression function. Fisher's formulation employed the discriminant variables as the independent regression variables and introduced a dichotomous dummy variable as the dependent variable identifying group membership. Since its introduction in the literature, use of the regression function as a device for computing the discriminant function has been made by many researchers and discussed by several writers (see [4]) The preference of the regression approach among psychologists and educators seems to be due to the existence of numerous computer programs for rapid computation and also to their familiarity with regression methods.

The purpose of this paper is to consider a method presented in multiple regression for variable selection or deletion for the two group discrimination problem. The actual circumstances under which the method may be used in a strict sense seem to be relatively unknown to researchers in educational fields.

Background and Notation

In order to specify more exactly the problem and notation, let us review very briefly the method of constructing the linear discriminant function, relate this result to the multiple regression function, and finally inspect the task of deleting or adding variables in this setting. We begin by specifying the notation for two random samples from two populations. Let X_{1jk}, $j = 1, 2, \ldots, J$; $k = 2, \ldots, n_1$, represent value of the jth variate for the kth individual in sample 1 and X_{2jk}: $j = 1, 2, \ldots, J$; $k = 1, 2, \ldots, n_2$, represent the jth variate for the kth individual in sample 2. We assume that we have two independent J-variate normal populations with means $\mu_{11}, \mu_{12}, \ldots, \mu_{1J}$ for the first population and $\mu_{21}, \mu_{22}, \ldots, \mu_{2J}$ for the second population and a variance-covariance matrix $[\sigma_{pq}]$, $p=1, 2, \ldots, J$; $q = 1, 2, \ldots, J$ common to both populations. The estimates of μ_{1j} and μ_{2j} are $\overline{X}_{1j.}$ and $\overline{X}_{2j.}$, the sample means for the jth variate for samples 1 and 2 respectively.

The estimate for σ_{pq} is:

$$s_{pq} = \frac{\sum_{i=1}^{2} \sum_{k=1}^{n_i} (X_{ipk} - \overline{X}_{ip.})(X_{iqk} - \overline{X}_{iq.})}{n_1 + n_2 - 2}$$

Under these assumptions Fisher's discriminant function has certain well-known optimal characteristics. By definition the linear discriminant function based on J variates is

$$(1) \quad L = c_1X_1 + c_2X_2 + \cdots + c_JX_J$$

and the difference between the means of the two sample discriminant functions is

$$\overline{D}(J) = \overline{L}_1 - \overline{L}_2 = c_1(\overline{X}_{11.} - \overline{X}_{21.}) + c_2(\overline{X}_{12.} - \overline{X}_{22.}) + \cdots + c_J(\overline{X}_{1J.} - \overline{X}_{2J.}).$$

Now if the ratio of the variance of L between the two samples to the variance of L within the two

samples is maximized, proportional solutions of the c's in (1) above are obtained. As a useful computational technique, Fisher [1, 184] has shown that if a dummy variable Y is introduced, where

$Y = n_2/(n_1+n_2)$ if a sample item is a member of sample 1

$Y = n_1/(n_1+n_2)$ if a sample item is a member of sample 2

then the calculation of an ordinary multiple regression equation will be equivalent to the linear discriminant function with the partial regression coefficients equivalent to the c's of (1). Of course, the problem remains unchanged if the Y are transformed so that Y equal 0 or 1.

Fisher [1,185] has also shown that, from the regression point of view $R^2(J)$, the square of the sample multiple correlation coefficient based on J variates, is

$$R^2(J) = \left(\frac{n_1 n_2}{n_1 + n_2}\right)\bar{D}(J) \Big/ \left[1+\left(\frac{n_1 n_2}{n_1 + n_2}\right)\bar{D}(J)\right]$$

In addition he specifies a test for the hypothesis $\rho(J) = 0$, (i.e., that the population multiple correlation coefficient based on J variates is null) to be made by means of $F = R^2(J) / J \div [1-R^2(J)] / (n_1 + n_2 - J - 1)$ which under the previously listed assumptions is distributed as the F-distribution with J and $n_1 + n_2 - J - 1$ degrees of freedom. Finally, using the same assumptions, this test can be shown [2, 378] to be identical to the test of the hypothesis that $\Delta^2(J)$, Mahalanobis' Generalized Distance based on J variates, is zero.

The Regression Method for Deleting Variables in the Discriminant Function

With the above material as motivation it seems reasonable to pose the following question. If the regression approach to the two group discriminant problem is taken, what process may be utilized so as to obtain a smaller (or larger) set which will discriminate well? In this paper we have not considered the more involved (and more important) problem of the total process of selecting the best set of variates for optimal discrimination, but only a part of this process.

In ordinary regression analysis the task of deleting (or adding) independent variates may be handled by means of an hypothesis test. An investigator may elect to reduce a total number of J variates down to J* variates if the multiple correlation coefficient is not significantly decreased as a result of such reduction. He, therefore, will be concerned with testing the hypothesis, $\rho(J) = \rho(J^*)$ where $\rho(J)$ and $\rho(J^*)$ are multiple correlation coefficients based on J, the original number of variates, and J*, the final number of variates after deletion of J-J* variates.

Under the assumptions of regression the test can be made by computing

$$(2)\quad F = \frac{(n_1 + n_2 - J - 1)}{(J - J^*)}\left[\frac{R^2(J) - R^2(J^*)}{1 - R^2(J)}\right],$$

which is distributed as the F-distribution with $J - J^*$ and $n_1 + n_2 - J - 1$ degrees of freedom. It is clear that the usual assumptions of multiple regression are not met in the present discriminant analysis. However, given the result in [2, 378], it would seem that the test using (2) might be closely related to a test on the Generalized Distance Function.

Accordingly, reference to Rao [3, 246] reveals that the Generalized Distance Function between two multivariate (J-variate) populations is defined as

$$\Delta^2(J) = \sum_{p=1}^{J} \sum_{q=1}^{J} \sigma^{pq}(\mu_{1p} - \mu_{2p})(\mu_{1q} - \mu_{2q})$$

where σ^{pq} is the pqth element of the inverse of the covariance matrix $[\sigma_{pq}]$ and $\mu_{1p} - \mu_{2p}$ is the difference between the two population means for the pth variate. Furthermore, $\Delta^2(J)$ can be estimated by

$$D^2(J) = \sum_{p=1}^{J} \sum_{q=1}^{J} S^{pq}(\bar{X}_{1p.} - \bar{X}_{2p.})(\bar{X}_{1q} - \bar{X}_{2q.})$$

where S^{pq} is an element of the inverse of $[S_{pq}]$, the sample estimate of $[\sigma_{pq}]$. Rao [3, 247] shows that Fisher's $\bar{D}(J)$ and Mahalanobis' $D^2(J)$ are related by the expression

$$\bar{D}(J) = \frac{D^2(J)}{n_1 + n_2 - 2}$$

and we know from Fisher [1, 185] that

$$R^2(J) = \frac{\left(\frac{n_1 n_2}{n_1 + n_2}\right)\bar{D}(J)}{1 + \left(\frac{n_1 n_2}{n_1 + n_2}\right)\bar{D}(J)}$$

so that

$$(3)\quad R^2(J) = \frac{\left(\frac{n_1 n_2}{n_1 + n_2}\right)\frac{D^2(J)}{n_1 + n_2}}{1 + \left(\frac{n_1 n_2}{n_1 + n_2}\right)\frac{D^2(J)}{n_1 + n_2 - 2}}$$

If in (2) we substitute for $R^2(J)$ its equal in (3) by means of algebra we obtain

$$(4)\quad F = \frac{[n_1 + n_2 - J - 1]}{J - J^*} \times \left(\frac{n_1 n_2 \,[D^2(J) - D^2(J^*)]}{(n_1 + n_2)(n_1 + n_2 - 2) + n_1 n_2 D^2(J^*)} \right)$$

This statistic is seen, by reference to Rao [13,253], to be the statistic which may be employed in testing the hypothesis, $\Delta^2(J) = \Delta^2(J^*)$, i.e., the Generalized Distance is the same when J variates are used for discrimination as it is after J-J* variates are eliminated so as to obtain J* variates. Again reference to Rao [3, 253] indicates that the F in (4) is distributed as the F-distribution with J-J* and $n_1 + n_2 - J - 1$ degrees of freedom under the assumptions listed before.

So it is seen that the test normally employed in regression analysis can be used here under, of course, differently specified assumptions. It also seems that it is rational to retain a set of variates for discriminating between the two populations if their deletion does not result in a significant drop in the generalized distance between the two samples.

It should be noted the expression (4) is of somewhat general form. For example, it may be desired to delete one variate at a time in which case J* is simply J-1 and the test is equivalent to testing the significance of a given discriminant coefficient. Moreover, it may be of interest to ascertain if discrimination (as measured by Generalized Distance) with J variates is essentially the same as that with no variates at all. In this last case J* = 0 and so we start with J variates and check the effect of deleting J variates. This test is, of course, equivalent to testing $\Delta^2(J) = \Delta^2(0) = 0$ as in (2).

Although the present test of significance may be used to advantage in a number of circumstances, it is true that in any practical case we run into trouble almost immediately. First, the debilitating effect of the assumption of multivariate normality and of common covariance matrix is somewhat unknown. In addition, the test will probably have to be applied many times in sequence such that any joint probability remarks made at the end of the sequence of tests will be difficult to obtain.

Summary

We have considered the use of the multiple regression technique of using a dummy dependent variate to obtain Fisher's linear discriminant function. A method of variable reduction was specified utilizing the test of $\rho(J) = \rho(J^*)$, i.e., the multiple correlation coefficient based on J variates is equal to the multiple correlation coefficient after the number of variates is reduced to J* ($\leq J$) variates. This test procedure was shown to be equivalent to testing $\Delta^2(J) = \Delta^2(J^*)$, under a proper change in assumptions, that the generalized distance between two J-variate populations is not affected by reducing the set of discriminant variates from J to J*.

REFERENCES

1. Fisher, R. A. "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, VII (1936) pp. 179-88.
2. Fisher, R. A. "The Statistical Utilization of Multiple Measurements," Annals of Eugenics, VIII (1938) pp. 376-86.
3. Rao, C. R. Advanced Statistical Methods in Biometric Research (New York: John Wiley, (1952).
4. Tatsuoka, M. M. and Tiedeman, D. V. "Discriminant Analysis," Review of Educational Research, XXIV (1954) pp. 402-20.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

A TRIANGULAR-DISTRIBUTION APPROXIMATION OF THE AREAS UNDER THE UNIT NORMAL CURVE

Y. LEON MAKSOUDIAN
University of Minnesota

IN MANY statistical problems there is a distributional assumption involving the use of the unit normal curve. Because the evaluation of this distribution function,

$$\int_{L'}^{L''} (2\pi)^{-1/2} \exp(-x^2/2)\, dx,$$

is rather complex, the usual procedure is to make use of the values already tabled. There are occasions, however, when an investigator would wish an approximation to the portions under the unit normal curve without having to refer to such tables. A simple method will be suggested in the following section which is a quick means of obtaining these portions without reference to tables.

Derivation of Method

Since the area under the unit normal curve is 1 square unit, and the ordinate at the mean is

$$(2\pi)^{-1/2} = 0.3989 \ ,$$

or approximately .4, let us consider an isosceles triangle with an altitude of 0.4 and base of 5 units as a density function. To demonstrate that the area of the triangle is a "good" approximation of the unit normal curve, we proceed as follows:—

Superimpose the triangle over the unit normal curve such that their means coincide and the base of the triangle coincides with the x-axis (see top of next page). Suppose it is desired to find the area labeled A under the normal curve, where L is any value between 0 and 2.5. The area of the trapezoidal segment OLba can be readily found from its geometric configuration, namely by solving for the area A in the relationship.

$$\frac{0.5 - A}{0.5} = \left[\frac{2.5 - L}{2.5}\right]^2 ,$$

which reduces to

$$A = \frac{1}{100}\left[8L(5 - L)\right] .$$

By simple differential calculus it can be shown that the inflection points of the unit normal curve are at ± 1 and the ordinate value at those points is 0.2420. The corresponding value for the triangular analogue is 0.24, which implies that the values obtained from the triangle will be underestimated for the portion from 0 to 1 and overestimated for most values over 1. If we were to assume that these error differences are uniform in L and therefore suggest a +0.001 correction to be added for each 0.05 increment up to L = 1.00 and -0.001 correction for each 0.05 for L over 1.00, we would obtain a closer fitting of the two distribution functions. A few computational examples will suffice to demonstrate the method.

Some Computed Examples

Example 1: Let L = 1

$$A = \frac{8}{100}\left[1(5 - 1)\right] = 0.32 \ .$$

Correction factor: $1/.05 = 20$; $20(+0.001) = 0.02$

Therefore, $A = 0.32 + 0.02 = 0.34$

Tabulated value: 0.3413

Example 2: Let L = 2

$$A = \frac{8}{100}\left[2(5 - 2)\right] = 0.48 \ .$$

Correction factor: $20(+0.001) + 20(-0.001) = 0.$

$A = 0.48 + 0 = 0.48$.

Tabulated value: 0.4772

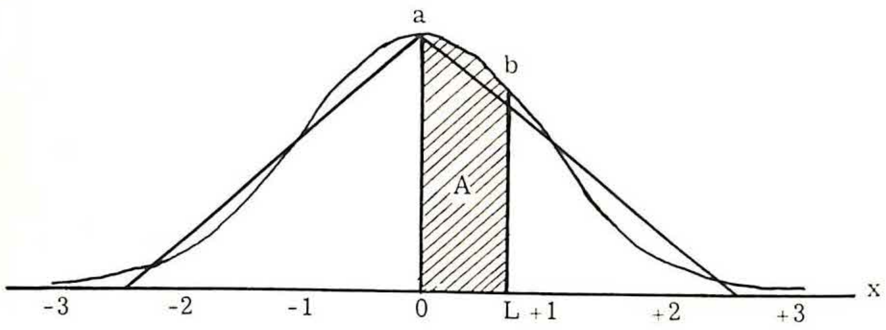
Example 3: Let L = 1.25

$$A = \frac{8}{100}\left[(1.25)(5 - 1.25)\right] = 0.3750 \ .$$

Correction factor: $20(+0.001) + 5(-0.001) = 0.015$.

$A = 0.3750 + 0.015 = 0.3900$.

Tabulated value: 0.3944



The list below shows the comparison for representative values of L, and indicates the error involved.

L	Method	Normal Table Values	Error
0. 25	. 1000	. 0987	+. 0013
0. 50	. 1900	. 1915	-. 0015
0. 75	. 2700	. 2734	-. 0034
1. 00	. 3400	. 3413	-. 0013
1. 25	. 3900	. 3944	-. 0044
1. 50	. 4300	. 4332	-. 0032
1. 75	. 4600	. 4599	+. 0001
2. 00	. 4800	. 4773	+. 0027
2. 25	. 4900	. 4878	+. 0022
2. 50	. 4900	. 4938	-. 0038

The technique to find the area between any two specified values of L' and L'', where the relation $-2.5 \leq L' < L'' \leq 2.5$ is satisfied, is identical to the usual method employed in computing the area from the unit normal curve.

Summary

A simple and easily employed method has been presented for estimating the areas under the unit normal distribution function by fitting an isosceles triangle. Examples have been provided which show the simplicity of the use of this method for the determination of the areas under the unit normal curve. The method outlined here can be used, when no normal curve probability tables are available, to yield quite close approximations to the actual values.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

A SIMPLIFIED METHOD OF MATRIX INVERSION

LOYAL W. JOOS
Baltimore County Schools
Towson, Maryland

STATISTICIANS AND others who frequently have the task of inverting large matrices usually make use of some method such as the Doolittle, in which linear operations are performed on the matrix and simultaneously on a unit matrix, so that the unit matrix becomes transformed to the desired inverse. These procedures have the disadvantage that a working area roughly twice that of the matrix is required. For desk calculations, the desired area can be provided by sufficiently large sheets of paper, but when the procedure is set up for computer operation, the provision of sufficient memory locations may be impossible for a large matrix. Naturally, it is precisely for such instances that the computer is most needed, because of the extreme labor required to invert large matrices by any other means.

This paper presents a procedure for inverting a matrix without the use of an appended unit matrix. All operations requiring fixed memory location are performed so that only those locations which were needed for the original matrix are required, provided only that a value may be erased and replaced by another in the same location.

In the following discussion, these definitions will apply:

1. Pivot Element — the matrix element of each row which lies on the principal diagonal of the matrix.
2. Pivot Row — the matrix row which includes the pivot element being used at that time.
3. Constant Multiplier — the element of each row which lies in the same column as the pivot element being used at the same time.
4. Constant Multiplier Row — the matrix row containing the constant multiplier in use.

The method of inversion to be described can be reduced to these five steps, which are repeated cyclically until the matrix is inverted:

1. Obtain the reciprocal of the pivot element and store it in that same location.
2. Multiply the reciprocal of step 1 by each other element in the pivot row and store each product in the same place from which the element was obtained for multiplication. This amounts to dividing every element in the pivot row by the pivot element, except for the pivot location.
3. Obtain the constant multiplier for a row and store it temporarily in memory. Zeroize the location from which the constant multiplier was obtained.
4. Multiply the constant multiplier by each element of the pivot row and subtract the product from the vertically corresponding element of the constant multiplier row.
5. Repeat steps 3 and 4 for each row of the matrix other than the pivot row. Repeat all five steps until every pivot element has been used.

A study of the example will reveal that the procedure is very similar to the Doolittle method, yet involves no appended unit matrix,. Since all calculated values are kept in their proper relationships, matrix inversion is easily seen to be a series of linear operations performed on the matrix itself, rather than on an appended matrix.

It is interesting to note that the several operational cycles can be performed in any sequence with identical results. Since most matrices encountered by researchers are symmetrical, the calculations are self-checking to some extent, in that the matrix is symmetrical at the end of each cycle, except for algebraic signs.

When these procedures are programmed for the computer, it will be found easier to multiply the reciprocal of the pivot element by every element in the pivot row, following which the reciprocal is stored in the proper location. This makes it possible to do step two identically each cycle, beginning multiplication with the first element in the pivot row.

Example:

Matrix to be inverted:

		2.50000	3.83333	4.95172
		3.83333	7.94361	5.43210
		4.95172	5.43210	9.71520
Cycle 1				
Steps one and two		.40000000	1.53333200	1.98068800
Steps three and four	(3.83333)	-1.53333200	2.06584244	-2.16053073
Steps three and four	(4.95172)	-1.98068800	-2.16053073	- .09261238
Cycle 2				
Steps three and four	(1.53333200)	1.53808631	- .74223085	3.58430056
Steps one and two		- .74223085	.48406402	-1.04583519
Steps three and four	(-2.16503073)	-3.58430056	1.04583519	-2.35217145
Cycle 3				
Steps three and four	(3.58430056)	-3.92376494	.85144026	1.52382624
Steps three and four	(-1.04583519)	.85144026	.01905912	- .44462541 = Inverse
Steps one and two		1.52382624	- .44462541	- .42513908

Note: Brackets above contain constant multiplier used with a line. Values underlined occupy pivot element locations. While the above matrix is small, it is sufficient to illustrate the procedures.

End of Special Section on Measurements, Statistics and Methods of Experimental Research, Edited by Drs. Collier and Stanley

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

THE SPREAD OF AN INNOVATION: HIGH SCHOOL LANGUAGE LABORATORIES[1]

RALPH NORMAN HABER
Yale University

WHEN ONE ASKS how new scientific or technological ideas spread from their creators to potential users, adequate answers are hard to find. There are a number of studies of diffusion of agricultural information, due in part to the efforts of the Federal Extension Service of the United States Department of Agriculture, which permit "field" studies in their natural settings. The majority of these reports, over the past thirty years, have appeared in Rural Sociology, and have investigated sociological variables almost exclusively.

In an entirely different direction, Coleman, Katz, and Menzel (1960) have studied four communities of doctors, specifically with respect to some of the variables underlying adoption of a new drug. The authors were interested in how knowledge of the innovation spread among doctors, the time-curve of adoption, the patterns of communication among doctors and its effects on adoption, how the results of use were evaluated, and factors leading to individual doctors' predispositions toward adoption.

These two areas of study, agriculture and drug-therapy, are illustrative examples of the spread of innovation in the natural sciences. Two recent books have reviewed a wide range of literature of diffusion of innovation, covering primarily natural science innovation, although between them, they have included studies from many other areas. Everett M. Rogers' The Diffusion of Innovations (1962) is the most general source on this subject, although the volume edited by D. H. Ross, Administration for Adaptability (1958) covers over 150 diffusion studies.

When the focus of attention is specifically on the behavioral sciences, several reasons are suggested to indicate that processes of innovation there will be different from those in the natural sciences. The behavioral sciences rarely can offer the same stringent criteria for evaluation of the effectiveness of the innovation. Hybrid corn clearly has greater yield per acre than older seeds, and simple experimentation will document this. But how does a potential adopter tell whether changes in child-rearing practices leads to improved development in his child. The reasons for this difference are complex, though not relevant here. But because of them, the adopter of behavioral science innovations is usually unsure, before he has made the investment, whether or not it will work.

A second reason is that there are non-existent or poorly organized patterns of communications regarding behavioral science knowledge. Fewer middlemen or distributors are selling, for example, techniques of mental health, child-rearing practices or occupational aptitude testing, compared to both of the natural science examples above.

Thirdly, every person is already an expert in the behavioral sciences. He can observe himself and others, and has formed stable ideas of what people are like, and of what their motivations are for doing things. Consequently, he will tend to reject any contrary information in these areas.

A related impediment to behavioral science innovation is that such changes are more apt to run counter to the established ways of doing things than are changes resulting from new natural science knowledge. If a new innovation is an addition rather than a substitution for established practices, it should be relatively easy to achieve adoption. Most technological innovations are of this nature. But the majority of behavioral science changes are replacing already existing ways of acting, in which there is often considerable personal investment. Hence, the difficulties in adopting these new innovations are immeasureably greater.

Each of these reasons indicate that dissimilar processes may be underlying the adoption of behavioral science innovations as compared to those in the natural or physical sciences. These reasons also help to specify some of the important decision-making conflicts that are present in any adoption situation. The present study was undertaken to see what these processes of adoption look like in one specific area of the behavioral sciences, that of changing the methods of teaching foreign languages in high schools. Its major purposes are to describe the processes, and to suggest hypotheses that may have some generalizability to

other innovation problems.

The spread of a new innovation in education was selected for study— the use of tape and disk recorders in language laboratories to teach foreign languages in high schools. The definition of what constitutes a language laboratory has a great range of application, but specifies as a core, some equipment enabling the student to listen to recorded drill reading exercises, usually with native voices, and permitting him then to repeat what he hears. Further, some laboratories have equipment to allow the student to record and listen to his own voice, giving him a direct comparison not possible without the recordings. Laboratories differ in the elaborateness of their equipment, and the number of tape units, listening booths, recording channels and students that may use the facilities simultaneously. The costs range from only several hundred dollars to $30,000., and the range in space is from only a few earphones and one recorder in the back of the regular classroom to a number of separate laboratory rooms used only for laboratory purposes.

The history of this innovation is brief, with widespread adoption resulting from army language training during the second World War. Its origins are hazy, but the first record of an established language laboratory is at Middlebury College in Vermont in 1915. By 1939, 33 colleges and universities had laboratories, and the most recent estimate, relative to the date of this study (Department of Health, Education and Welfare, January, 1959), puts the total at 240 colleges or universities, or about 13% of the nearly nineteen hundred in the United States.

In high schools, which are the concern of this study, 70 language laboratories had been established by 1958, according to the same survey by H. E. W., assessing how many schools had established laboratories. However, since the definition of what constitutes a laboratory was left to each school, there may be a few who do have laboratories by the standards specified above, but not by their own, and hence did not report this information to H. E. W. In any case, the number of schools with language laboratories is less than one-half percent of the 21,000 high schools in the United States. Thus, while this innovation has been available since 1943 in high schools (and since 1915 in colleges and universities) it has, as yet, been adopted by only a few schools. This implies that all of the schools in the present sample should be treated as early adopters, and that the big spurt in the adoption of this innovation has not yet occurred.

Procedure

For the study, an open-ended questionnaire was mailed to each of the high schools known to have a laboratory. It was addressed to the chairman of the language department and was answered by that chairman in all but one case. Thus, the information was obtained from the actual users of the innovation, rather than from merely administrative officials.

The records of H. E. W. as of January, 1959, indicated 63 high schools had laboratories, according to each high school's own definition. In addition, seven schools with laboratories, not included in the H. E W. records, were added to the sample, bringing the total known high school laboratories to 70. The initial questionnaire was mailed in the late Spring and Summer of 1958 to 63 schools, including these seven not listed by H. E. W., and omitting seven which were listed by them. These latter seven were mailed in January 1959. Answers were received from 62, or 89% of the 70 known laboratories. Of those 62, one was unable to answer the questions because its laboratory was established prior to the appointment of its current staff; two have not established their laboratory yet; one had recently switched to disk recorders (and hence did not feel they should answer the questions- no response on a further inquiry); and seven are still experimenting with their laboratories, and were unable to answer the questionnaire. Therefore completed and applicable responses were received from 51 schools, which will comprise the sample for this study.

The results will be presented question by question. Numbers enclosed by parentheses refer to the number of laboratories falling within the category specified by the text. All analyses of relationships are tested against the chi-square distribution corrected for continuity (McNemar, 1962), with the two-tailed probabilities reported. When it is meaningful to refer to responses to questions before or well after they are discussed, a parenthetical reference will show the question number, the value of chi-square, and the probability of incorrectly rejecting the null hypothesis. There are a few occasions when the number of schools is less than 51 for some comparisons, due to omission of an answer.

No control sample was selected for this study, that is, schools who had not adopted this innovation, either with or without exposure to it. The decision to exclude a control group was reached on the grounds of expediency, but its effects on the interpretation of the study are important. All of the comparisons to be reported are within this sample of 51 adopters. Whenever the adopters are homogeneous with respect to some variable, then that variable can provide no information as to how it relates to any other in the study, or as to how it relates to adoption of behavioral science knowledge. A further limitation on interpretation stems from considering all of these schools as early adopters. Thus, not only are they not compared to non-adopters, but also not to late-adopters.

The function of this study is not primarily just

to examine how these 51 schools did establish their laboratories. More importantly, it is attempting to analyze the antecedents and processes of innovation in general, drawing hypotheses from this particular case study. Three separate questions are involved: what conditions lead to exposure of the school to the idea; what qualities of the school predisposed it to consider the idea more favorably; and what kinds of processes are involved in reaching a final decision to adopt. Because of the inherent limitations of mailed questionnaires and with the lack of a control sample, very little information will be available on the second question. Both the agricultural studies, and the drug studies do have material relevant to predispositional variables. However, the present study does provide data on both the communication and the process questions, and these will be examined in detail.

Before turning to the results, several general statements relevant to these questions may help to create a framework for the analyses, and for some suggested hypotheses.

Looking first at the exposure and communication aspects of innovation, we are concerned as to who in the school initially heard of the innovation, what were the sources of information, and what channels were used, who in the school spread the idea to others, and what further sources of information were available and sought.

Regarding the processes leading to adoption; what were the stages of investigation of the feasibility of the idea, what kinds of opposition developed, and from what sources and for what reasons, what kinds of support were present, and from what sources; what was the pattern of decision-making; and what kinds of arguments were used to mobilize further support.

The questionnaire posed sixteen questions. Before considering each one separately, the total list is presented here. Since some of the analyses are considered out of order, before they are discussed fully, this list can serve as a guide to all of the questions. The order of presentation in the paper has been altered slightly from that of the questionnaire for ease of discussion.

Q 1. When was the language laboratory installed?
Q 2. How long did the school take to install a laboratory after the idea was first raised?
Q 3. Who brought up the idea first?
Q 4. Where did the initiator hear about the laboratory?
Q 5. Who were the prime movers?
Q 6. Who made the final decision regarding the adoption of the laboratory?
Q 7. Who did the school check with for further information?
Q 8. What kinds of information were obtained in this checking?
Q 9. From what sources did the idea receive support?
Q 10. If there was any opposition, where did it come from?
Q 11. What were the greatest problems in establishing the laboratory?
Q 12. Were there any arguments raised against the idea of the laboratory itself?
Q 13. Did the school set up a pilot study before the laboratory was fully established?
Q 14. What convinced the school of the need for the laboratory?
Q 15. Has the school communicated with others regarding the laboratory?
Q 16. What is the current opinion of the laboratory now?

Results

Q 1. When was the language laboratory installed? The oldest high school language laboratory dates from 1943, although two-thirds of them (34) began in the last three years prior to the study (1956-1959). This indicates that all of these schools should be classified as early adopters, responding at a time when clear proof of the feasibility of the laboratory has not been fully tested. However, one of the important questions of this study regards the factors determining the date of adoption, and the decision processes leading up to it. Consequently, lacking a control group makes it necessary to divide these schools into earlier and later "early" adopters, and examining the differences between the two groups with respect to these decision processes. As the relevant variables are discussed, these differences will be spelled out.

Q 2. How long did it take the school to install a laboratory after the idea was first raised? The difference in time between when the idea was first raised and the final installation is an estimate of the time-lag of adoption of this innovation. One-third of the schools (16) had a delay of less than one year, while another third took only one or two years(17). The longest lag was ten years. While it is difficult to compare times for different innovations, this appears to be a short lag compared to the agricultural studies.

There was no relationship between length of time-lag and the date of installation (Q 1) – the earlier laboratories did not take longer to install. However, schools with long time-lags were more likely to have experimented with a pilot study first (Q 13: $X^2 = 4.5$, $p < .03$). Since a pilot study takes a minimum of one year, and in most cases several years, before the results can be evaluated and an adoption decision made, it is not surprising that schools that conducted pilot experiments took longer to install their laboratory.

Q 3. Who brought up the idea first? In over two-thirds of the schools (35), a language teacher was the first instigator of the idea. In addition,

in eight schools, the instigators were principals, in four they were supervisors, and in four schools, other persons introduced the idea. While these findings are not startling, they are important in evaluating the growth of the laboratories. The major source of agitation for the laboratories comes from inside, from the people who are primarily responsible for foreign language teaching, from those who are going to use the laboratories.

Q 4. Where did the initiator hear about the laboratory? There are three major sources of information: 1) attendance in college courses that taught or used laboratory facilities (16), 2) visits to other high schools (4) or college labs (14), and 3) professional literature (20). In addition, previous teaching experience in a high school with a laboratory (6), and personal experimentation (3) contributed to sources of information. These account for all but eight schools, six of whom were unable to answer the question because the origins of the laboratory were too remote. Thus, 24 of the initiators have had direct experience with language laboratories before, while 21 knew of them through visits or literature. Therefore, it is not enough to say that language teachers account for the instigation of this innovation; more specifically, they are those language teachers who received modern teaching training, and who are motivated to keep up with new developments in their field.

Since there are no schools in which the initiator has not had any contact at all with laboratories or literature about them before, it is only possible to compare differences between those who have had direct experience with laboratories and those whose knowledge is attained indirectly. If the initiator is a teacher, he is more likely to have had direct experience (Q 3: $X^2 = 3.5$, $p < .06$). There is also a slight tendency for those schools whose initiators had direct contact to have fewer problems regarding use of the laboratory (Q 11: $X^2 = 2.9$, $p < .10$). But contrary to this, there is clearer evidence that schools whose initiators had direct contact are <u>more</u> likely to conduct pilot experiments (Q 13: $X^2 = 5.6$, $p < .02$). Thus, if the initiator is a language teacher or he knows about the laboratories as a result of his own training and teaching, the school is more likely, not less, to do further experimentation. It is the schools that already have a backlog of personal knowledge of how to use the laboratories that want more. It is almost as if those that know how to use it also know that it needs extensive testing. Those who have never used it rush in where the proverbial angels fear to tread.

Q 5. Who were the prime movers? This question asked who was primarily responsible for carrying the idea through to adoption. Not surprisingly, thirty-nine schools reported that a language teacher was the prime mover. Twelve of these same schools also mentioned their principal as a prime mover. Twelve schools mentioned some other administrative official. The audio-visual department (4), parents (3), outside foundations (3), college language teachers (2) and students (1) were also mentioned as prime movers. There were only twelve schools in which a language teacher was not the prime mover. However, of the 39 schools where they were, 20 indicated that some other person was also instrumental, 17 of whom were principals or supervisors. This undoubtedly reflects the power and decision-making hierarchy in the schools, since, according to Q 6, no language teacher was in a position to make the final decision regarding adoption.

The relation between the initiator and the prime mover was very high. Of the 36 schools in which language teachers were initiators, 33 of them also mentioned teachers as prime movers (Q 3: $X^2 = 50.0$, $p < .001$). Thus asking who was the initiator and who was the prime mover elicits the same information, indicating that initiators continue to carry the brunt of the arguing, persuading and convincing. Whether this is a necessary pattern cannot be ascertained without a control sample of non-adopters, but it seems like a reasonable possibility.

Q 6. Who made the final decision regarding the adoption of the laboratory? The principal did for 19 schools, the school board for 17, the superintendent for 8, and seven schools merely said administration, which may be referring to any of the other categories. In no case was the language teacher or language chairman able to make this decision. Thus, while the majority of the initiators were teachers, they needed support from some administrative official before adoption could occur (see Q 5). There are no relationships between the official who makes the decision to adopt and any of the other variables in the study. Certainly, it would seem that the closer and more accessible the administrative power is to the language department (i.e., principal vs. school board), the easier it would be to reach a decision favorable to the department. However, if the principal, who is a closer source of power, is against the innovation, he is probably more destructive to the innovation than a more diverse though distant board. In any event, whatever differences there might be do not show up in this study. Perhaps with a control group of non-adopting schools, the pattern found here would be more unique.

Q 7. Who did the school check with for further information? Three major sources of information were used by the schools: 1) college language laboratories, either from discussion with college language teachers, or from courses in college or in post-graduate workshops (33); 2) equipment firms who advised the schools regarding the types of electronic equipment available (35); 3) other high schools with already existing laboratories (22). In addition, schools checked with audio-visual experts (6),

equipment firm catalogues only (5) and professional literature only (3). Four schools indicated that their own staff had all of the requisite knowledge, and hence did not check with anyone.

There was no relationship between the sources that the school checked with and the date of origin of the laboratory. But there is a slight tendency for those schools who checked with colleges, or whose teachers took post-graduate work, to have longer time-lags (Q 2: $X^2 = 2.2$, $p < .15$). Thus, of the 16 schools that did not check directly with a college laboratory program, only four of them took two years or more to install their laboratory. This seems to indicate that acquiring more experience takes more time, and and seems to slow up the process of innovation. Whether it makes the final decision more likely to be positive, however, cannot be determined from this study, since all of these schools eventually decided to adopt the laboratory. There is also a relationship between where the initiator acquired his information and where the school checked further regarding the laboratory, so that if the initiator had direct contact with laboratories, the school was more likely to turn to college laboratories and workshops for more information (Q 4: $X^2 = 6.0$, $p < .02$). Of the 23 initiators who had direct contact, 20 of them returned to the colleges for their information and training.

Q 8. What kinds of information were obtained in the checking? Two major problem areas were discussed by the schools with their informants. One was the type of equipment to use in the laboratory, with emphasis on problems of cost, installation, space and maintenance (34). The other was how to use the laboratory, with special regard to techniques of teaching, scheduling of classes, making of tapes or disks, student reaction, and the like (24). Most schools indicated that both kinds of information were important, and hence, sought (22). However, two schools checked only on teaching information while 12 checked only on equipment. In addition, there were the four schools who did not check with anyone, and 13 who indicated they checked, but did not get much information, either because their own teachers knew how to use the laboratory already (7), or because this was so new that no one knew anything about it at all (6). Nine of these 13 did indicate they obtained some other kind of information, while only 4 of the 13 said they got nothing. Six schools were unable to answer the question. Initiators with direct laboratory training tended to seek further teaching rather than equipment information in their checking (Q 4: $X^2 = 2.9$, $p < .10$), which serves as further support for the finding that those who know want to know more. In addition, the schools that turned to colleges for more information about the laboratories, obtained information primarily about teaching techniques (Q 7: $X^2 = 2.8$, $p < .10$). There is, however, no relationship between the kinds of problems the school had with the laboratory (particularly teaching problems) and the information they obtained (Q 11: $X^2 = p > .05$). Thus, it did not seem to be specific information that sent the schools to check with college sources, but rather that they were already familiar with that source, due to the previous training of their staffs.

However, there is one exception to the conclusion that initiators with direct laboratory training tend to seek further teaching information, for when the initiator is a language teacher, teaching information is not sought as much as would be expected. It is the non-language-teacher-initiator who obtained the information regarding teaching techniques (Q 3: $X^2 = 5.3$, $p < .03$). The non-teachers needed teaching information more, as indicated by the relationship between teaching problems and non-teacher-initiator (Q 3-11: $X^2 = 2.8$, $p. < 10$) Thus, while the school's problems with the laboratory do not determine the information sought, both teachers familiar with college laboratories, and administrators who need information to solve technique problems, seek teaching technique infor[illegible]tion; the former, because they know a ready source, and the latter, because they are ignora[illegible] of the solution. If one looks at non-teacher-initiators in schools with teaching problems, they sought teaching technique information almost exclusively (Q 3, 11-8: $X^2 = 6.8$, $p < 01$).

Teaching information was sought more by the more recently established laboratories (Q 1: $X^2 = 3.0$, $p < .08$), which are also the schools indicating they had more problems with teaching techniques for the laboratory (see Q 11).

Q 9: From what sources did the idea receive support? Here, also it was the efforts of the language teachers, as reported by nearly half of the schools (20), which provided the main support and encouragement for the innovation during the process of installing the laboratory. However, both the principal (15) and the city school administrators (18) were also mentioned as strongly in support of the innovation. Several schools mentioned, in addition, support from parents and community organizations (7), non-language teachers in the school, particularly from the audio-visual department (6), students (3), and outside foundations (2). Four schools indicated that there was no appreciable support (even thought the laboratory was established)'.

Looking at the relationship between supporters and initiators, support came mainly from teachers when a non-teacher was the initiator (Q 3: $X^2 = 2.9$, $p < .09$), though teacher initiation did not necessarily elicit support from the administration (Q 3: $X^2 = 0.5$, $p > .50$). There was more support from the teachers in the laboratories that were established earliest (Q 1: $X^2 = 3.4$, $p < .07$), (though the later ones did not receive more from administrative officials), and those schools where the support came primarily from teachers, further teaching information was most likely to be sought

(Q 8: $X^2 = 5.3$, $p < .03$).

Q 10. If there was opposition, where did it come from? Half of the schools (24) indicated that there was no opposition at all. This must have been because so many schools were able to install their laboratories very quickly. Schools with no opposition had shorter time-lags (Q 2: $X^2 = 8.0$, $p < .01$). Of the sources of opposition mentioned by the other schools, nine said the school administration was not behind the project, five said the language teachers did not back it, and four simply mentioned that the school was too old-fashioned. Of course, all of these schools were able to overcome whatever opposition there was, since by the definition of the sample, all of them have language laboratories.

Schools established later (1957-59) had less opposition to contend with (Q 1: $X^2 = 3.2$, $p < .07$). This may be due to an awareness by the later schools of the success experiences of other laboratories. While there is no relationship between the amount of opposition and who was the initiator (Q 3), there is one with the source of the initiator's idea, there being less opposition if he had direct contact with laboratories before (Q 4: $X^2 = 4.4$, $p < .05$). Thus, opposition is based, not so much on who the initiator is, but rather how much specific knowledge he has about laboratories.

If an administrative official was providing support, there was much less opposition to the innovation (Q 9: $X^2 = 7.4$, $p < .01$). This would seem to imply that if the administration was not behind the innovation, they were against it. There was, however, only a slightly shorter time-lag when an administrative official supported the innovation (Q 2: $X^2 = 1.1$, $p < .40$), indicating that administrative support did not speed the adoption of the laboratory appreciably. When teachers rather than administrators supported the laboratory, the time-lag was substantially shorter (Q 2: $X^2 = 8.0$, $p < 01$). Hence it seems that teacher support is a far more critical factor in determining the time-lag.

Q11. What were the greatest problems in establishing the laboratory? There were several specific problems that problems that most schools had to solve regarding their laboratories. In order of mention these were: costs (33), space (21), techniques of use (17), scheduling (16), selection and maintenance of equipment (10), and teacher motivation (9). Only three schools said they had no problem at all. Of these categories the interesting ones were the techniques of use and teacher motivation, for these referred not to the physical and financial obligations of innovation, but to knowledge and willingness to use it. Most schools would be expected to have financial problems arising from the demands of new facilities—two-thirds of this sample specifically mentioned them. While these problems might be grounds for opposition to the laboratory, once the funds were made available, the opposition should disappear. However, when the problems regard knowledge of how to use the new facilities, or a willingness to do so, then less reason exists to expect that this type of opposition will disappear so readily. With this expectation, analyses were based upon differences between schools who mentioned only mechanical problems – cost, space, and the like – and those who mentioned teaching problems.

Those schools who mentioned any kind of teaching problem had more opposition to the laboratory (Q 10: $X^2 = 12.6$, $p < .005$). Teaching problems were more prevalent in the more recently established laboratories (Q 2: $X^2 = 4.0$, $p < .05$), and also in schools where the initiator was a non-teacher (Q 3: $X^2 = 2.5$, $p < .12$). There was also a slight tendency for more teaching problems to be found when the initiator's source of the innovation was based on direct contact with laboratories previously (Q 4: $X^2 = 2.1$, $p < .15$). Strangely, however, there were no relationships between the nature of the problems presented by the innovation and whether a pilot study was undertaken (Q 13), nor with the sources or amounts of further information sought (Q 7 and Q 8), nor with the length of the time-lag (Q 2).

Thus, with these findings, it is clear that the kind of problems posed by the new laboratory were quite important in the interaction with other factors regarding adoption. While the presence of teaching problems did not slow down the speed of adoption, it did seem to be the basis for much of the opposition to the laboratory (which did increase the time-lag (Q 10). The teaching problems seem to occur mostly in schools where the laboratory is recent. This would seem to imply that these kinds of problems can be worked out with time. But apparently, the presence of teaching problems before adoption does not prevent adoption. In fact, there is no relationship between any kind of problem and the length of the time lag. Apparently, the time lag is not affected by problems, only by opposition or support.

Teaching problems also seem to occur where a non-teacher was the initiator. There is little evidence to show how this might specifically create teaching problems, since it was previously shown that the language teachers supported the laboratory more when a non-language teacher was the initiator (see Q 9). It seems most reasonable to explain this as an idea being introduced before the staff was fully prepared to utilize it.

Q 12: Were there any arguments raised against the idea of the laboratory itself? Eighteen schools indicated that some arguments had been raised (or are still being raised) as to the value of this technique of teaching or to the idea of any change or innovation. These kinds of objections to the laboratory seemed to be quite different in nature to those based upon lack of money or space, or even to those based upon lack of knowledge about techniques of use. However, when they are

compared with these other objections, a pattern of great similarity is found, so that there may be little independence.

Teaching problems were found only in schools where objections to the innovation were raised (Q 11: $X^2 = 31.0$, $p < .001$), though there were a few schools where objections to the innovation were present and where there were no teaching problems. Amount of opposition and objections to innovation are highly correlated (Q 10: $X^2 = 10.6$, $p < .01$).

There are several instances of the objections variable relating uniquely to other factors. For example, while teaching problems did not increase the time-lag (Q 2), objections did (Q 2: $X^2 = 4.6$, $p < .04$). Objections were also more prevalent in later laboratories (Q 1: $X^2 = 2.4$, $p < .12$), though so were teaching problems.

It is unfortunate that more information is not available on this question, but questionnaires are difficult instruments with which to elicit adequate responses in potentially explosive areas. It is also more likely that the value of this question would show up more if a control sample had been used. Objections to this innovation, or to innovating in general may be one of the major reasons why more schools have not adopted language laboratories as yet.

Q 13: Did the school set up a pilot study before the laboratory was fully established? Over two-thirds (35) of the schools did have a pilot study of some kind, while 16 did not. Remembering that all of these schools currently do have laboratories, these latter 16 adopted this new technique without any formal testing.

Previously, it has been indicated that pilot studies were conducted in schools with longer time-lags (Q 2: $X^2 = 4.5$, $p < .05$); though there was no relation between pilot studies and opposition (Q 10), nor between pilot studies and date of installation (Q 1). Also discussed previously was the finding that if the initiator knew about laboratories as a result of his own training or through having taught in a school that already had a laboratory, then the school was more likely to have a pilot study (Q 4: $X^2 = 5.6$, $p < .02$).

There was no relation between conducting a pilot study and sources of support (Q 9), nor with whom the school checked for further information (Q 7), nor with the kinds of information obtained (Q 8). This latter finding is interesting because apparently the decision to conduct a pilot study does not rest on specific knowledge obtained elsewhere, or on specific problems presented by the laboratory (Q 11).

Since whether a pilot study is conducted or not requires a major decision by the school, it is important to examine the kinds of factors that seem to underlie it. There are only two clearly significant findings from the questionnaire. Those schools who run pilot studies have a longer time-lag, and have initiators who got the idea for the laboratory from their own training and teaching. None of the other variables seem to relate in any productive manner. Particularly, there are no relationships between this decision and the kinds of problems the laboratory had, or with the way the school set about to solve them.

It would seem that the pilot study caused the longer time-lag rather than the other way around, so that leaves only the initiator's source of inspiration as a possible causal factor. This was previously discussed as perhaps being due to the fact that those initiators who have already been exposed to the laboratory knew some of its problems and wanted to test it out more fully before commitment. This is still a weak argument, and without much other substantiation in these data.

Here again, a control sample is needed to clearly evaluate the importance of pilot testing. Two-thirds of the schools in this sample did test before installation, which is a large percentage. There may be many other schools now in the process of testing, but as yet do not have a full laboratory. It would also be expected that the earlier schools would test more, since they do not have the benefits of already established laboratories and staffs with experience. The failure to find a relation between date of installation and pilot studies is probably due to the fact that all of these schools are early adopters. The 16 schools in this sample who did not experiment were not homogeneous enough for this questionnaire to show how they were able to avoid doing so.

Q 14: What convinced the school of the need for the laboratory? The one major reason mentioned was that the laboratory allowed better ways of teaching foreign languages. There were many elaborations of this idea, but nearly three-fourths of the schools (40) indicated that the specific improvement in the methods of teaching was what convinced the administration.

There were other important answers given too. The laboratory would save time for teachers (10), outside pressures were applied (11), particularly concerning better language preparation for college entrance requirements (this reason is perhaps the same as for the first one given, except that in the former 10 cases, the schools innovated on their own, while with these eleven, they recognized the value of the innovation, but did not themselves innovate until forced to do so.);other schools have been successful with their laboratories (6); it is inexpensive (6) (sic); the laboratory is flexible and can be used by other programs in the schools (7); students like it (5); it is one aspect of a generally progressive program (5), and there is a great need today to know more languages (4).

There are three categories of response that could be classified as rationalization for adoption—the laboratory really saves teacher's time (10), the laboratory is flexible and can be used by other programs (7), and external pressures were applied

to get adoption (11). At least one of these reasons was mentioned by 26 schools. They were classified as rationalizations in that the school did not adopt the innovation solely as a means to lead to improvement in the teaching of foreign languages.

Schools that used these rationalizations were distinguished from the others in two cases; they had more teaching problems (Q 11: $X^2 = 8.0$, $p < .01$), and they were established later (Q 1: $X^2 = 3.0$, $p < .08$). Since a school that had teaching problems would have trouble reporting that the laboratory was a better way to teach, the first finding is understandable. And since the later schools had more opposition in general (Q 10), these rationalizations may have been one of the ways to overcome some of that opposition.

Q 15: Has the school communicated with others regarding the laboratory? Fifteen schools had presented their experiences and results at meetings or conventions, while 41 reported that other schools and individuals (in addition to the author of this case study) have corresponded with them regarding their laboratory. Only five schools said they had not talked to other schools about their laboratory after it was established. Comparing the 15 schools who have presented their experiences and results at professional conventions or meetings with those who have not, those laboratories that presented were established earlier (Q 1: $X^2 = 2.2$, $p < .15$). This is probably due mainly to the greater length of time they have had available to do so, though it may also imply either that the earlier schools were more conscious of the needs for their knowledge to be shared, or that they have experienced greater pressures on them to share what they have found out.

Further, schools that had presented their experiences had initiators with indirect sources of knowledge about the laboratory (Q 4: $X^2 = 2.2$, $p < .15$), had less support from the language teachers (Q 9: $X^2 = 2.6$, $p < .11$), and had sought less teaching information in the process of establishment (Q 8: $X^2 = 3.6$, $p < .06$). These findings listed last are quite puzzling, since they do not seem to make sense. Communication of one's experiences to colleagues and other professional persons would seem to be an indication of both knowledge about the innovation, and the kind of interest in the innovation that would lead to proselytizing and encouraging others to try it out. These last three findings do not seem to describe either the person or the school that would be interested to do this.

Q 16: What is the current opinion of the laboratory now? Over two-thirds of the schools (36) said it was favorably or enthusiastically viewed by everyone. Fourteen schools indicated that the laboratory needed improvement, mostly in learning how to use if effectively. One school said the laboratory was not essential, and one school abandoned their laboratory due to a shortage of space and teachers to operate it.

The only relationship found was between early opposition to the innovation and the current opinion of the laboratory now, where there was a slight favoring of good opinions in schools where there was no initial opposition (Q 10: $X^2 = 2.0$, $p < .15$). There are no other relationships that even approach significance. This may reflect upon the form of the questions asked – i. e., in answer to the questionnaire, schools expressed a more positive view than they really have – or that, more likely, there is not enough variablility among this group of schools.

Implications

This completes the presentation of the results of this study, question by question. It is now necessary to begin to examine their implications in terms of the general characteristics of the innovation process, rather than just those concerning high school language laboratories.

In the introduction, two aspects of innovation were discussed as relevant to this particular study–communication or antecedent variables, and decision or process variables. A third aspect, predisposition variables, were excluded from the scope of this project. This final session will discuss the first two, and by summarizing the findings regarding high school language laboratory adoption, it will attempt to draw some general hypotheses about behavioral science innovation.

Certain qualifications need to be made at the outset. These schools are all early adopters. The time course for this innovation has merely begun. Hence, even if the questionnaire used was a perfect instrument, much of the needed information would still be unavailable. Further, most of the early laboratories had a fairly short time lag between the time of first discussion and of final adoption–short by comparison with other adoption situations. Consequently, all of the process variables are telescoped, and patterns are hard to discern. The implication of this short lag, at least by inference from the agricultural studies, is that as the innovation spreads, the lag will become even shorter, reaching perhaps some kind of asymptote, representing the decision-making time within the school. For this sample, there was no evidence of this–the laboratories established later did not take less time. This is further evidence for the homogeneity of the "early" adopters in the sample.

Perhaps the clearest and most recurrent finding throughout the analyses was the prominence of the language teachers in the adoption of the language laboratories. Moreover, it was specifically those language teachers with modern training or those who have kept up with current improvements in the professional language publications who were most responsible for the initiation of the idea in the school. And of those two categories of

sources, modern training was more important than merely being well-informed. Thus, the initiator was most likely to be a teacher who had himself been trained with a language laboratory in college, or who had taken post-graduate training with such a laboratory. In sum, the innovation was not brought in from outside, but rather was initiated by those who were familiar with it before, and by those who would be using it in the schools. The middle-man functions did not seem to play a large part here at all, at least not in the initiation stages. In both the agricultural and the drug therapy studies, the ideas were introduced from the outside by highly specialized middlemen. This difference will be explored in more detail below.

A second characteristic of the teacher-initiator was his quest for more knowledge and training with the language laboratory. Schools in which the initiator was well trained sought even more information and training. There seems to be several aspects to this. One is clearly that initiators, regardless of who they are, turn to whatever source they are familiar with, when they need further information. Thus, there is evidence that regardless of the problems posed by the laboratory, the school seeks further information from previously explored sources. Their checking is not problem-oriented, but source-oriented, and this source orientation is based almost exclusively on their own previous experience. The second aspect is that those initiators who are well-trained or experienced with the laboratory are just the ones who return for more complete information and further training. It is as if they realize how much more there is to be known about using a language laboratory. For example, it is in schools where the initiators have had direct contact with a laboratory that pilot studies are more likely to be done, rather than the reverse. Likewise, if the teacher was a strong supporter of the laboratory, teaching information was more likely to be sought. Schools where the initiator had more knowledge about laboratories also reported more problems in using the laboratory. All of these instances clearly point to the conclusion that those who know more, know there is more to know. Or perhaps this is further support for the old adage — a little knowledge is a dangerous thing — and those with the most knowledge recognize this and try to acquire more.

Whether this result is due to the fact that knowledge of a good source is possessed only by those who used it before or due to the fact that some knowledge breeds the need for more, cannot be clearly specified here, though undoubtedly both processes are involved.

When we look at the pattern of expedition and support, the importance of the language teachers becomes even more evident. As a general conclusion, adoption of this innovation requires, at all stages, the support and active participation of the teachers who are going to use it. Without this, prospects for adoption seem nil, though of course, since there are no schools in this sample without laboratories, this assertion cannot be proved at the moment. But the pattern within this sample is striking. The initiators are language teachers, the prime movers or expeditors are language teachers, the source of further information is determined by previous sources of the language teachers, the enthusiasm of language teachers leads to a shorter time lag, which is not true for purely administrative support, and finally, there are fewer teaching and technical problems when a language teacher initiates the idea of the laboratory within the school. The implication of all of these findings is that this innovation, to be adopted, requires specifically the initiative support and approval of those persons who will be directly concerned with its use. This is true, even though for every school in the sample, the final decision rested with an administrative officer.

In spite of an expectation that most expediting would be done by the administration, the language teachers apparently played an important role here. However, this is not to deny the power of the administration, since of the 39 schools where a language teacher was a prime mover, 17 reported that an administrative officer was also a prime mover. But the language teacher-initiator continues to carry the brunt of the arguing, persuading, and convincing. This is probably a different pattern than in most innovation situations, where the initiator has no responsibility beyond the introduction of the idea. For example, suggestion boxes, used as techniques for seeking new ideas, clearly divorce the source of the idea from its further implementation.

What appears to be happening here is a spread of the innovation within the profession of language teachers, with little information about it outside of those persons. For the high schools, the impetus seems to be coming from the recent college-trained teachers who were exposed to language laboratories during their training, or who have returned to college workshops and refresher courses to keep up with new developments. It is clear that the college and university laboratories are the major source of the innovation. Most of the questionnaires ultimately point back to a college laboratory as the original source. While 20 schools mentioned professional literature as a source, there were only 10 who said it was an original and only source. Two schools indicated the idea originated with the initiator. Thus, only 12 schools, less than 25% of the sample, lacked contact with a language laboratory before initiation.

There are several important implications of this dependence on language laboratories, particularly college language laboratories. It should be expected that the adoption rate of this innovation will grow rapidly, as more laboratories are established in both high schools and colleges. Each one can serve as a potential stimulus for many others.

This is particularly true of the college laboratories, where language teachers-in-training are exposed to this method of instruction, and are given specific training in its use. While this study gathered no data on the source of each initiator's training, it is clear that many of them were trained in the newer college laboratories.

The reverse side of this is also interesting. There is greater difficulty in adoption if the iniator is a non-language teacher. This implies that this innovation not only does come from inside, but seemingly must come from inside. This might be a general characteristic of behavioral science innovation. Because these innovation problems represent ways of doing and thinking that individuals already possess, and hence, must change, it is difficult or impossible to impose these changes. The specific individuals concerned must be ready for them – in fact, they must want to change themselves.

A further implication is that this innovation is implanted primarily in training, and is much less likely to be picked up at a later time (though one-fourth of the sample here did). This is perhaps similar to other instances of behavioral science innovation, where initial exposure or training is most important, and latter attempts at retraining are poor substitutes. If it is not stretching the analogy here, trying to change child-rearing and child care practices run up against difficulties when the parent was taught differently or raised differently. If would seem therefore, that if a change is to be made, it would be facilitated if a trained person is implanted into the system. Some innovations can clearly be accepted by the potential user even without previous training. Others meet more resistance if they run counter to a current practice, technique or way of thinking. One of the characteristics of many behavioral science innovations is this particular difficulty. This reasoning explains, perhaps, the absence of middlemen in the spread of this innovation. They would be relatively ineffective without the pre-trained recipients with which to start. Another expectation is that, while 75% of this sample had direct contact with laboratories before, the percentage of indirect innovations will increase with the later adoptions. As the laboratories become more established and accepted, schools will be more willing to accept them on their face value, without the direct contact and testing. Here, middlemen will certainly have more effect. It should be noted that middlemen did begin to function for these schools in the evaluation process of the innovation. Two-thirds of the schools reported that they checked with equipment firms later on. However, for these schools, this was not the path for the introduction of the idea. That may change in future years.

The length of the time lag was influenced by several variables. Most important was whether a pilot study was undertaken, and further by the amount of checking that was done with others. Clearly, acquiring further information and experience takes time, though it may be that those schools that checked or experimented are by nature more cautious and would have taken longer in any event.

In addition to these variables, the time-lag was influenced by the amount of opposition and the number of objections to the innovation. If there were none of these, little time was lost in adopting. This was particularly true of support came from the language teachers. This is all reasonable and to be expected. The large number of schools who reported no opposition, objections or special problems may help to account for the short time lags. Why there should be few of these problems will be discussed below.

Aspects of support and opposition have already been mentioned, but since they represent an important set of conditions in the processes of adoption, some further elaboration is necessary. Several different questions were asked regarding these problems: sources of support, sources of opposition, reasons for opposition, objections to the idea of the innovation, and finally, problems in the establishment of the laboratory. There is an overall correlation among these, due primarily to the fairly large proportion of the schools that indicated that they had support from everyone, no serious problems to be overcome, and no opposition or objections to the innovation. The specific positive relationships have been summarized in the respective sections of the results of the questionnaire. Again, it is important to keep in mind that all of these schools adopted the innovation – hence, none of them could have experienced insurmountable opposition or they would not have been in the sample. Undoubtedly, if these findings were compared with schools who have been exposed to the idea, but have not adopted, the amount of opposition would be a larger factor. Even so, within this sample, there are interesting patterns. For example, the opposition does affect the time lag, lengthening it, as would be expected. Further, the opposition seems to be a function of the source of the initiator's information – direct previous contacts with laboratories arouses less opposition, regardless of who the initiator was. Thus, it would seem that opposition is based upon the source of the information, not upon the persons involved. There was less opposition when the administration was an active supporter of the laboratory – implying that if the administration was not behind the idea, they were opposed to it. If teachers supported the laboratory, a shorter time lag was achieved, though support of the administration had no effect of shortening the time lag. Administration support was necessary, but clearly, it was not sufficient. Schools that had teaching problems had more opposition, probably because schools with teaching problems had less direct contact with laboratories before. The

presence of opposition flavored current evaluation of how the laboratory was working out, making them appear less positive.

One third of the schools had objections to the innovation itself, and these increased the time lag. All of the schools reporting teaching problems also reported, in addition, some of these objections, indicating that in large part, the objections were influenced by difficulties or uncertainties regarding the use of the laboratory, rather than any physical or financial problems with it. Teaching problems also seem to be important with regard to the reasons or rationalizations given by the school for the adoption – schools that used predominantly rationalizations had more teaching problems, and hence, presumably, could not point to the reality of how well they could teach with the laboratory.

Except for the question regarding objection to the innovation itself, all of the support and opposition questions were answered in concrete terms of the process and problems of adoption. One would have expected more personal biases to show up than did. Undoubtedly, this lack was due to the questionnaire method. We did find, when we talked personally with language teachers in several of the schools in the sample, that the feeling of individuals were very much involved. For example, few of the questionnaires mentioned problems in convincing older teachers that the laboratory was a good technique, even though this was a central issue in each of the personal interviews. Undoubtedly this was reflected in the "objections" question, but in a somewhat disguised form. However, the expectation that this kind of innovation would be difficult to adopt because it runs counter to established ways of teaching foreign language was not found. Perhaps, our results are really correct in that these schools may have been predisposed to try new ideas without any traditional-directed hold-outs on their faculty. However, it is reasonable to expect that this must be a problem and will continue to be a problem in later adoptions.

When we looked at the decision-making directly, it was difficult to discern any clear-cut patterns. In every school an administration officer made the final decision. No language teacher or even a language department chairman was capable of doing this. Beyond this, however, there was little uniformity. Among the 51 schools, all of the various administrative officers were represented as being able to make the final decision. Some schools apparently were able to decide to adopt only several months after the first official mention of the idea occurred. Others spent many years in discussions, explorations, training, testing and checking before a formal decision to adopt was made. Pressure, in general, was applied by some or all of the language teachers on the administration, or in some cases, by an alliance of the language teachers and some administrative officers on higher administrative bodies. There was no uniformity of hurdles that had to be met. Some schools demanded very complete and proven justification before approval was given. Others were willing to go ahead on its potential only.

Already discussed was the necessity of support from the language teachers. This is perhaps the strongest aspect of the decision-making pattern. The administration did not shape policy here; they merely approved it. This, in large part, may represent the reality of any large organization like a school system, and if so, it does not reflect upon any special qualities of behavioral science innovations. However, one would expect that if this were merely a decision about physical plant, or some other non-behavioral science problem, that the administration would not be so restricted by the attitudes of the members of their staffs. In other words, it is possible that when an innovation is a behavioral science one, i.e., one that involves changes in current practices on the individual level, the administration cannot dictate policy. They must be led from below, or they must ateempt to re-educate those for whom the adoption will represent a change.

FOOTNOTE

1. This project was undertaken while the author was a research associate in the Institute of Communication Research at Stanford University, under the direction of Professor Wilber Schramm, during 1957-58.

REFERENCES

1. Coleman, J., Katz, E., and Menzel, H. Doctors and New Drugs: A Case Study in the Diffusion of an Innovation. (New York: The Free Press. 1960).
2. McNemar, Q. Psychological Statistics. 3rd Edition. (New York: Wiley, 1962).
3. Rogers, E. M. The Diffusion of Innovation. (New York: The Free Press, 1962).
4. Ross, D. H. (ed.) Administration for Adaptability. (New York: Teachers College, Columbia University (New York Metropolitan School Study Council, 1958).
5. United States Department of Health, Education and Welfare. Foreign Language Laboratories in Schools and Colleges. Bulletin 1959, No. 3. Complied by M. C. Johnston and C. C. Seerley. 1959.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

A STUDY OF SOME INFLUENCING FACTORS UPON AND THE NATURE OF YOUNG CHILDREN'S WRITTEN LANGUAGE

LOUIS ADA WILSON
Wisconsin State College
Superior, Wisconsin

THERE ARE certain characteristics of children's written compositions and influencing factors which, when analyzed, will produce findings that should indicate procedures for teaching written language.

The Problem

To analyze the recorded words from the written language lessons from a number of different groups of children in the elementary school in order to ascertain, if possible, the factors that influence the writing of these lessons and determine the characteristics prevalent.

The Need and Purpose

After having made a study of children's needs in spelling, the writer was aware of the need for teachers to know the nature of children's writings and the factors influencing their written compositions.

The interest was in finding the cause for the seeming need to escape for some children when they were asked to write the language lesson: The tendency to delay, to move about, and to engage in conversation indicated a need for release from restraints that were not obvious within the situation.

Assumptions Necessary in Understanding the Study

A. The procedure used was kept constant through the period of the study but certain elements in the situation were changed as each group came or before the group arrived in the third grade. The decisions were made after a contemplative review of the accomplishment of each group of children. These decisions always came from a conscientious effort to help children to write more and better language lessons.

B. Throughout the experiment the children were taught written language using interests and needs as a basis for selecting the activities in the here and now situation in the integrated classroom procedures; units of study were carried on; culmination of these units were dramati ation, radio and literary programs, exhibits and parties with the children entertaining pupils from other classrooms, college students, and parents of the children. These were always to celebrate the children's success in learning. There were included picnics, field trips, and visits from interesting persons who talked with and to the children.

C. There was an abundance of literature for Social Studies and Reading Lessons. The children had an improvised movable stage. There were maps, and a globe, and also, a set of encyclopedias. The art room supplied paints and crayons as needed for teaching and learning. Science was also a vital part of the children's experiences. Music, singing, group piano lessons, plays and games were interesting experiences and a happy part of the children's school day.

D. The Wisconsin course of study was checked carefully to estimate that children were getting English skills comparable to those of other children in the state.

The philosophy had once been tried that permitted the children to write without emphasis on the mechanics of English. It had been successful as far as the ideas presented were concerned, yet the children's writings were so full of errors in mechanics of writing, this philosophy was discarded.

Scope and Delimitations

In teaching the written language lessons with

the different Third Grade groups, the variations were as follows:

A. These groups, I, II, III, IV, V, were taught one full year.

In Group I, the philosophy that children will write when they have something they want to write was followed.

In Group II, all children were required to write during the written language period.

In Group III, after reviewing the teaching and Group II's written language papers, the decision was made to still continue requiring the children to write, meanwhile the teacher in second grade was asked to continue using manuscript writing throughout the year. Previously, the change to the cursive handwriting was made during the last half of the second grade, so the change was then made at the beginning of the third grade.

In Group IV, the cursive writing was delayed until the last half of the third year.

In Group V, manuscript writing was used throughout the year. (Permission was given by the Director of Teacher Training to permit the child to continue using manuscript writing th gh the third grade.)

At the close of this year, the writer went back to graduate school for a year of study, and continued the study upon returning to the teaching situation.

B. These three groups, VI, VII, and VIII, were taught one-half year each.

Children in Group VI were taught the first semester of the school year, the teacher teaching them all their subjects. They used cursive writing, having made the transfer from manuscript in the second grade.

In Group VII, the children were using cursive writing. The language lesson was lengthened from 25 minutes to 40 minutes, since the writer taught only the written language lessons for the second semester of the school year. Children were changed from manuscript writing the first half of the third-grade year.

In Group VIII, the children were using cursive writing. They had made the change the first half of the third-grade year. The language lesson was 25 minutes long daily for the second semester of the school year, the writer teaching only the written language lesson.

The Method of Analysis Employed

The written language was analyzed for:

A. The number of running and different grammatical usages of nouns, verbs, adverbs, adjectives, pronouns, prepositions, articles, infinitives, conjunctions, and contractions.

B. The number of mechanics of English usages as to sentence length, number of sentences, periods, capitals, commas, questions, apostrophes, abbreviations, titles, and compositions.

C. Statistical calculations were made for:
1. correlations between chronological age and the number of different nouns, verbs, adverbs, adjectives, pronouns, prepositions, sentence length, number of sentences, periods, and capitals.
2. correlations between the IQ's and the same phases of English listed in 1 above.
3. correlations between frequency of different usages, between nouns and adverbs, between verbs and adverbs, between adverbs and adjectives, and between nouns and verbs.
4. the mean and standard deviation (sigma) of chronological ages, intelligence quotients, the number of different nouns, verbs, adverbs, adjectives, pronouns, prepositions, sentence length, number of sentences, capitals, and periods.

Findings

The numerical counts of the different grammatical and different mechanics of English usages show:

Group V using manuscript writing throughout the third grade had the greatest number of usages in all areas studied (see Table VI), except in use of contractions. The use of contractions occurred most often in Groups VII and VIII who had been changed to cursive writing in the first half of the third grade. Group VII used contractions more frequently than Group VIII, but Group VIII had a 25-minute language lesson while Group VII had 40-minute lessons.

Group V also had the highest numerical count of words and usages for the number of running usages except for use of infinitives and commas.

Groups II, III, IV, V, and VI had been taught that contractions were to be avoided. This point was not equally stressed with Groups VII and VIII where another teacher taught the subjects other than the language lesson.

These numerical counts showed a steady increase in the number of words written by each group as the use of manuscript handwriting was extended each half year. For example, from the Groups: II, changed to cursive handwriting in

second grade, used a total of 2,623 words; III, changed to cursive writing in the first of the third grade, used a total of 3,798 words; IV, changed to cursive writing in the last half of the third grade, used a total of 4,636 words; V, which continued manuscript writing throughout the third grade, used a total of 7,681 words.

Groups VII and VIII were changed to cursive writing in the first half of the third grade. Group VII had a 40-minute language lesson while Group VIII had a 25-minute language lesson. The total words used by each group were: VII, 5,920 words, and VIII, 3,647 words, indicating the influence of time upon the amount of writing done during the language lesson.

The numerical count indicates that the parts of speech for each group ranked as follows, nouns, verbs, adjectives, adverbs, pronouns, and prepositions. The frequencies follow the rank order as noted by Brandenburg (1), McCarthy (2), and Nice (3), whose studies were of spoken language of preschool children. This is shown in Table VI. Thus, the findings indicate that children's written language shows a pattern of usage similar to spoken language of an earlier age.

The count of the running words indicated the order of usage to be nouns, verbs, pronouns, adjectives, and adverbs. The count of the running words for prepositions, articles, infinitives, and conjunctions showed a different ranking of these usages for each group. Table VI illustrates this.

The numerical counts for mastery in the mechanics of English show:

Group V to lead in number of usages except in the use of the comma.

Group VII used 607 commas while Group V used only 252. See Table VII.

Group VII had the tendency to use the comma with words in a series. They seemed to delight, whenever possible, to practice the skill learned in their written language. They had the knack of being loquacious and delighted in using words in a series. Also, it will be noted in Table VII that Groups VII and VIII used more abbreviations than Group V. Here again this had not been emphasized with Groups II, III, IV, V, and VI. Children in these groups had been taught it was better writing to spell the word out completely.

The numerical count of the pronouns shows that these children were linquistically immature in their written work. According to Goodenough, "the use of an unusually large number of pronouns after the age of three and one-half years may be interpreted as evidence of immaturity." (4) The written language showed that the number of running usages in all groups of pronouns had outranked both the adjectives and adverbs. However, the number of different pronouns used ranks fifth, i.e., the rank of different grammatical usages was then nouns, verbs, adjectives, adverbs, and pronouns. See Tables VI and VIII.

The analysis of each group's written papers indicated that the last half of the year tended to show the measure of success particularly as to the increase in learning the use of adjectives and adverbs which were taught progressively through each semester. In Group VI, for the first half of the year, the data showed a correlation of .57 between verbs and adverbs and of .40 between adverbs and adjectives, while Groups II, III, IV, V, VII, and VIII, for verbs and adverbs showed correlations of .87, .91, .96, .94, .83, and .78, respectively, and between adverbs and adjectives were respectively .64, .74, .93, .91, .71, and .71. These correlations indicate also that children using a number of different verbs will use relatively an equal number of different adverbs, and show as well that an equal relationship will exist between the number of different adverbs and different adjectives. These correlations ranging from .64 to .96 show moderate to very high relationship of usages. See Table V.

The correlations also indicated high relationship between other different usages, i.e., for each group between nouns and adjectives Group II .79, Group III .69, Group IV .90, Group V .86, Group VI .76, Group VII .81, and Group VIII .81; between nouns and verbs for each group they were .73, .91, .88, .92, .87, .89, and .72, respectively.

For those children, Group IV, starting cursive writing in the last half of the third grade and those in Group V using manuscript writing during the entire school year, there was only a slight difference in all calculations. Group IV had a median IQ of 115, while Group V had a median IQ of 110. Also, children in Group VII who had a median IQ of 113, had used cursive writing in the first half of the third grade and were given a 40-minute language lesson instead of a 25-minute lesson, showed only slightly lower scores comparable to Group V. When the IQ's are reviewed for the three groups, IV, V, and VII, the question arises: How much more would Groups IV and VII have written had they continued with manuscript writing through the third grade? Group II with a median IQ of 116 changed to cursive writing in second grade and had the least counts in all areas measured. These data indicate that the higher intelligence will not always overcome physical immaturity.

The relationship of chronological age with English usages was in general negative, except for Groups VII and VIII, who showed a low positive correlation for all items analyzed except for sentence length. These correlations ranged from .17 through .45 with -.05 for sentence length. Group IV in which correlations between IQ's and English usages and mechanics of English were high, the correlations between chronological age and usage

were negative, except for sentence length, .14, and .00 for prepositions. There needs to be further study here, particularly as to the rate of learning of certain skills by the brighter children and with larger groups of children.

The mean and standard deviation, sigma, were calculated for each group as to chronological age and intelligence quotient and as to the number of different English usages, such as nouns, verbs, adverbs, adjectives, pronouns, prepositions, sentence length, number of sentences, periods, and capitals. See Table II.

These calculations showed:

1. A range of mean from 5.0 to 5.9 with a sigma range of 1.1 to 1.6 for all groups as to sentence length, which indicates that two-thirds of the children in this third grade used about the same sentence length.

2. Group V, mean IQ 109.5, sigma 11.4, had highest mean as well as the largest standard deviation for all items analyzed, which indicated wide variables within the group. Yet this group showed the highest amount of written language during 25-minute language lessons using manuscript writing, and exceeded those children having 40-minute language lessons.

3. Groups V and VII used more pronouns than did other groups. Group V was younger, mean 101.7 months, in age while Group VII had the higher mean age, 113.5 months, sigma 11.8. Yet the difference of the means for each group in the use of pronouns was not great enough to be of significance. The excess use of pronouns indicates immaturity of speech (4) for Groups V and VIII and for all groups of this grade level.

4. The calculations of the mean indicate in general that children in grade three using manuscript handwriting write more and use the skills learned more than those using cursive writing. The amount of usage increases by half-years as the cursive writing skill is delayed. The sigma of the mean for each group as to the language usages and the mechanics of English signify wide variations within each group. The wider variations of the means and sigmas for the groups appeared in Group V using manuscript writing during the entire school year. See Table II.

Some of the subjective interpretations for this study were:

1. As the manuscript handwriting was extended with each group of children, there was less restlessness and less talking among them, especially in Groups III and IV.

2. When cursive writing skill was begun, the restlessness and talking increased in Groups III and IV. In Groups II, VI, VII and VIII, the children were restless and talkative with each other and with the teacher during the entire periods of the study.

3. Children in Groups VII and VIII seemed to want to please the teacher and others with their writings, while Group V was more concerned with what they wanted to write and how they wanted to say it.

4. From the experience with these six groups of children it is known that planned teaching experiences are essential in appropriate situations for learning the use of different adjectives, adverbs, and verbs. Extending the vocabulary for nouns comes within each unit of study.

5. There is need for further study as to the value of the subjective thinking in item four above in order to determine the long-time value of such planned learning situations for children.

6. There should be a definite policy among the teachers of the first three grades in teaching manuscript handwriting. This should be that children learn manuscript with the kinesthetic movement for each letter of the alphabet that is comparable to the cursive writing. Then the transfer from manuscript to cursive can be made easily. Some children have done this in as short a time as three weeks with the interpretation for them as to the manner in which the kinesthetic movement has already been learned.

7. When children are using cursive writing, they ask for the spelling of a great many words.

8. During the first-half of the third year when children are learning the cursive writing skill, language papers are seldom completed and often children do not wish to complete them.

Conclusions

The data show the following:

1. The delay of the cursive writing skill increases the mastery of written language each half year it is delayed. Group II changed to cursive writing in the second grade and wrote 2,623 different words, Group III changed to cursive writing in the first half of third grade and wrote 3,798 different words; Group IV changed to cursive writing in the second half of third grade and wrote 4,636 different words; and Group V which used manuscript writing during the entire year in third grade, used 7,681 words. This progress is also noted in the number of running words written.

2. The length of time spent in the language lesson influences the amount of written language. Children in Group VIII, having a 40-minute language lesson and using cursive writing in the first half of the third grade exceeded comparable Groups III and VIII who used cursive writing also in the first half of the third grade, and had a 25-

minute language lesson. The total numerical counts for different grammatical usages were: for Group VII, 5,920; Group VIII, 13,647; and Group III, 3,798.

3. The amount of written language produced by children seems to be correlated with intelligence. The correlation for each group with the intelligence quotients show a positive correlation. See Tables I and III. These data indicated that the higher the intellectual capacity the more children write. The change from manuscript to cursive writing in the second grade, Group II, indicates an influence on the amount of written language. In Group II, with a median IQ of 116 and a mean of 100.7 with a sigma of 13.2, the correlation between intelligence and for items analyzed ranged from -.23 through .20 may indicate that cursive writing is not for children at this age.

4. Children within a group will show a wide variation in their use of English and its mechanics. The mean and standard deviation, sigma, for the group that had the greatest amount of written language were as follows for the different usages: mean nouns, 121.7, sigma, 64.6; mean verbs, 77.9, sigma, 35.7; mean adverbs, 28.2, sigma, 13.9; mean adjectives, 47.8, sigma, 23.9; mean pronouns, 17, sigma 12.0; mean prepositions, 12.2, sigma, 4.3; mean sentence length, 5.5, sigma, 1.3; mean number of sentences, 97.3, sigma, 43.3; mean capitals, 149.9, sigma, 80.5.

5. Children using a large number of one grammatical usage use relatively an equal number in another classification. The correlation indicated the number of different nouns used will be related to the number of different adverbs used and different nouns to different adjectives, as well as the number of different adjectives used will be equivalent to the number of different adverbs.

6. Children in the third grade will, in their written language, show a pattern of speech that is similar to their oral speech used at an earlier age. The data show this to be true in the number of running grammatical usages. The rank order is nouns, verbs, pronouns, adjectives, adverbs, and prepositions. See Tables VII and VIII.

7. Children in third grade will in general, about two-thirds of them, use sentences of five words; for all six groups, the range of mean sentence length was 5.0 through 5.9, their sigmas ranged from 1.1 through 1.6.

REFERENCES

1. Brandenburg, G.C. "The Language of a Three-Year-Old Child," Journal of Genetic Psychology, XXII (1915), pp. 89-120.
2. McCarthy, D. Language Development in Children (Minneapolis, Minnesota: University of Minnesota, 1930).
3. Nice, M.M. "Concerning All Day Conversation," Journal of Genetic Psychology, XXVII (1920), pp. 166-77.
4. Goodenough, Florence. "A Note on the Development of Self-Awareness," Journal of Genetic Psychology, LII (1938), pp. 333-46.

TABLE I

THE INTELLIGENCE QUOTIENTS OF THE SEVEN GROUPS OF CHILDREN

		Year Groups			Half-Year Groups		
	Group II	Group III	Group IV	Group V	Group VI	Group VII	Group VIII
1	130	128	138	132	133	129	141
2	124	120	131	127	121	129	118
3	123	119	120	126	114	129	117
4	122	118	120	123	112	123	116
5	121	117	119	116	110	123	115
6	119	115	119	115	110	122	112
7	118	112	118	114	108	120	112
8	117	110	117	114	108	120	110
9	117	109	115	114	107	119	110
10	116	108	115	111	107	116	108
11	113	107	114	111	107	114	107
12	111	107	112	110	107	112	106
13	107	105	111	110	105	110	103
14	106	102	110	109	105	109	103
15	104	96	108	109	105	108	102
16	103	89	107	108	105	108	100
17	92	86	100	108	104	106	100
18	86	84	99	104	103	101	95
19	68		83	104	101	98	95
20			68	99	99	88	90
21				98	96	88	90
22				93	93		88
23				92	90		
24				81			
Change to Cursive	2nd Grade	1st Half of 3rd Grade	2nd Half of 3rd Grade	Continued Manuscript through 3rd Grade	2nd Half of 3rd Grade	1st Half of 3rd Grade	1st Half of 3rd Grade

TABLE II

THE MEAN AND SIGMA OF CHRONOLOGICAL AGE, INTELLIGENCE QUOTIENTS AND PHASES OF WRITTEN LANGUAGE OF THIRD-GRADE CHILDREN

No. of Children	19		18		20		24		23		21		22	
Groups	II		III		IV		V		VI		VII		VIII	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
C. A., and Mos.	100.7	13.2	101.0	6.9	101.0	8.0	101.7	7.5	100.0	4.3	104.5	4.4	106.4	5.2
IQ	110.4	14.7	107.3	11.8	111.2	14.9	109.5	11.4	106.7	8.6	113.0	11.8	106.3	11.6
Diff. Nouns	47.4	22.6	76.7	35.5	74.8	27.3	121.7	64.4	50.7	18.5	105.5	38.8	59.1	13.7
Diff. Verbs	32.8	15.0	48.0	16.3	58.0	19.4	77.9	35.7	35.6	13.1	66.4	25.7	37.5	9.5
Diff. Adverbs	14.5	8.2	17.7	9.2	23.0	9.8	28.2	13.7	15.6	4.9	25.6	9.3	17.5	7.0
Diff. Adj.	16.7	8.8	34.2	15.3	38.8	15.3	47.8	23.9	18.4	7.8	36.0	15.3	20.1	8.3
Diff. Pron.	9.8	4.5	14.0	3.4	14.5	2.6	17.0	12.0	12.8	3.8	18.1	5.6	14.5	3.0
Diff. Prep.	7.8	2.3	9.8	2.9	10.9	2.5	12.2	4.3	9.3	3.1	11.1	2.5	8.8	2.9
Sentence Length	5.0	1.6	5.1	1.3	5.0	1.4	5.5	1.3	5.9	1.5	5.1	1.1	5.0	1.5
No. Sentences	45.8	31.4	60.8	26.3	75.3	30.1	97.3	43.3	34.91	17.5	88.0	36.9	45.0	16.6
Periods	26.8	21.2	52.4	30.3	55.0	22.8	72.7	52.6	23.5	17.8	71.1	30.3	41.4	15.9
Capitals	104.7	70.3	102.7	55.0	131.2	54.7	149.9	80.5	54.4	33.1	141.8	59.0	84.8	30.4

Sigma represents the variability within the group.
Mean represents the general level of concept.
σ = Sigma.

TABLE III

THE CORRELATIONS BETWEEN THE INTELLIGENCE QUOTIENTS AND CERTAIN ASPECTS OF THE WRITTEN LANGUAGE OF SEVEN DIFFERENT GROUPS OF THIRD-GRADE CHILDREN

Length of Year	Four, Whole Year Groups				Three, One-Semester Groups		
	Group II No. = 19	Group III No. = 18	Group IV No. = 20	Group V No. = 24	Group VI No. = 23	Group VII No. = 21	Group VIII No. = 22
IQ and Diff. Nouns	-.02	.41	.45	.25	.56	.17	.10
IQ and Diff. Verbs	.04	.33	.49	.18	.57	.37	.16
IQ and Diff. Adv.	-.02	.40	.53	.17	.29	.36	.07
IQ and Diff. Adj.	-.23	.54	.49	.30	.50	.28	-.02
IQ and Diff. Pron.	.16	.25	.57	.24	.46	.52	.02
IQ and Diff. Prep.	.05	.13	.06	.15	.23	.46	.29
IQ and Sentence Length	0.00	-.26	.33	-.13	-.07	.21	.12
IQ and No. Sentences	.20	.49	.60	.39	.66	.25	.09
IQ and Periods	.10	.46	.71	.45	.51	.24	.14
IQ and Capitals	.08	.38	.61	.15	.70	.27	.16

TABLE IV

THE CORRELATIONS BETWEEN THE CHRONOLOGICAL AGE AND CERTAIN ASPECTS OF THE WRITTEN LANGUAGE OF SEVEN DIFFERENT GROUPS OF THIRD-GRADE CHILDREN

Length of Year	Four, Whole Year Groups				Three, One-Semester Groups		
	Group II No. = 19	Group III No. = 18	Group IV No. = 20	Group V No. = 24	Group VI No. = 23	Group VII No. = 21	Group VIII No. = 22
C. A. and Diff. Nouns	-.23	-.15	-.36	-.06	-.12	.27	.21
C. A. and Diff. Verbs	-.21	-.09	-.34	.02	-.06	.28	.21
C. A. and Diff. Adv.	-.14	-.18	-.38	.10	.11	.38	.22
C. A. and Diff. Adj.	-.04	-.22	-.34	-.02	-.07	.45	.35
C. A. and Diff. Pron.	-.32	-.17	-.49	.29	-.29	.17	-.02
C. A. and Diff. Prep.	-.31	-.15	0.00	.16	.09	.28	.36
C. A. and Sentence Length	-.18	.16	.14	.33	-.12	-.05	.32
C. A. and Sentences	-.20	-.16	-.35	-.06	-.13	.37	.12
C. A. and Periods	-.18	-.11	-.43	-.39	.09	.40	.13
C. A. and Capitals	-.17	-.12	-.31	-.04	-.02	.37	.17

TABLE V

THE CORRELATIONS BETWEEN DIFFERENT GRAMMATICAL FORMS FOUND IN THE WRITTEN LANGUAGE OF SEVEN DIFFERENT GROUPS OF THIRD-GRADE CHILDREN

	Four, Whole Year Groups				Three, One-Semester Groups		
	Group II No. = 19	Group III No. = 18	Group IV No. = 20	Group V No. = 24	Group VI No. = 23	Group VII No. = 21	Group VIII No. = 22
Different Nouns and Adjectives	.79	.69	.90	.86	.76	.81	.81
Different Verbs and Adverbs	.87	.91	.96	.94	.57	.83	.78
Different Adverbs and Adjectives	.64	.74	.93	.91	.40	.71	.71
Different Nouns and Verbs	.73	.91	.88	.92	.87	.89	.72

TABLE VI

TOTAL NUMBER OF DIFFERENT GRAMMATICAL USAGES IN WRITTEN LANGUAGE OF SEVEN GROUPS OF THIRD-GRADE CHILDREN

Groups	Nouns*	Verbs	Adverbs	Adjectives	Pronouns	Prepositions	Infinitives	Conjunctions	Contractions	Totals
II	901	623	275	317	187	149	67	69	32	2,623
III	1,380	864	319	615	252	177	106	76	6	3,798
IV	1,496	1,159	459	776	290	218	122	95	18	4,636
V	2,922	1,870	677	1,148	408	292	200	121	40	7,681
VI	1,165	819	358	424	295	215	74	63	31	3,447
VII	2,215	1,394	537	755	380	233	193	120	90	5,920
VIII	1,300	824	386	426	320	194	81	70	43	3,647
	11,379	7,553	3,011	4,461	2,132	1,478	843	614	260	31,752

* Proper nouns were counted even though of local significance.

TABLE VII

NUMBER OF USES OF THE MECHANICS OF ENGLISH MADE IN WRITTEN LANGUAGE OF EACH OF SEVEN GROUPS OF THIRD-GRADE CHILDREN

Group	Range of Sentence Length	Number Sentences	Periods	Capitals	Commas	Questions	Apostrophes	Abbreviations	Titles	Compositions
II	1 to 41	871	509	1,990	134	65	40	12	61	77
III	2 to 50	1,094	943	1,848	216	58	16	43	106	220
IV	1 to 34	1,505	1,099	2,624	131	103	17	29	182	278
V	1 to 44	2,335	1,745	3,598	252	127	58	48	234	500
VI	2 to 44	799	541	1,251	55	16	31	44	128	191
VII	1 to 68	1,847	1,494	2,978	607	72	48	72	141	374
VIII	1 to 32	990	911	1,865	287	57	29	85	105	230
		9,441	7,242	16,154	1,682	498	239	385	957	1,870

TABLE VIII

TOTAL NUMBER OF RUNNING GRAMMATICAL USAGES MADE IN WRITTEN LANGUAGE BY EACH OF SEVEN GROUPS OF THIRD-GRADE CHILDREN

Groups	Nouns*	Verbs	Adverbs	Adjectives	Pronouns	Prepositions	Articles	Infinitives	Conjunctions	Contractions	Totals
II	2,056	1,324	569	530	797	434	494	98	288	45	6,635
III	2,240	1,799	608	982	1,214	719	978	128	360	6	9,034
IV	2,580	2,484	830	1,272	1,700	783	985	154	481	24	11,293
V	4,959	3,898	1,456	1,968	2,489	1,495	1,913	284	1,044	59	19,565
VI	1,813	1,418	683	691	934	497	846	101	299	36	7,318
VII	3,437	3,418	1,229	1,338	2,742	963	1,154	327	964	207	15,779
VIII	1,862	1,567	689	644	1,222	487	711	110	331	55	7,678
	18,947	15,908	6,064	7,425	11,098	5,378	7,081	1,202	3,767	432	77,302

* Proper nouns were counted even though of local significance.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 3, Summer 1963)

AN EXPLORATORY STUDY OF THE ATTITUDES OF NONCRIPPLED CHILDREN TOWARD CRIPPLED CHILDREN IN THREE SELECTED ELEMENTARY SCHOOLS

HELEN K. BILLINGS
The University of Kansas City

Purpose of the Study

THIS STUDY WAS concerned with the problem of discovering what attitudes are held by noncrippled elementary children of Grades 1, 3, and 6 toward crippled children. It was an effort to find whether the factors influencing the formation of certain stereotypes such as race, nationality and religion, tend to become the same factors influencing the formation of stereotypes or prejudices regarding less "culturally important" characteristics, such as physical or mental deficiencies. It was primarily exploratory and descriptive in character.

Theoretical Position

The major theoretical consideration upon which this study was based was proposed by Roger C. Barker (1:8):

> Individuals who differ physically from the majority of individuals around them have the same minority status as minority group members, and as such, are subject to the dynamics that hold for minority groups... The minority status of the physically disabled, which is due to the negative attitudes of the physically normal majority, is, in almost all respects, similar to the problem of racial and religious, underprivileged minorities, although it may well be that the source of the negative attitudes toward the physically disabled is even deeper and less rational.

Hypotheses

The study investigated three major hypotheses:

Hypothesis I — The attitudes of noncrippled children toward crippled children tend to be different from their attitudes toward noncrippled children and these attitudes tend to be "unfavorable."

Hypothesis II -- There is a significant increase in the "unfavorableness" of attitudes toward cripples between grades 1 and 6.

Hypothesis III — The degree of "unfavorableness" of attitude toward cripples is related to the degree of satisfactory social-emotional adjustment of the school child.

Definition of Terms

Certain key terms were used in the study. They were defined in the following way:

Favorable attitudes are those which tend to result in verbal expressions and acts toward crippled children that tend to give them a feeling of acceptance, security, and worth, as measured by judges' inferences from the content material contained in the data procured through the two projective techniques.

Unfavorable attitudes are those which result in behavior toward the crippled child which tend to give him a feeling of rejection, insecurity, and worthlessness.

Neutral attitudes are those attitudes which cannot be said, in the opinion of the judges, to reveal either favorable or unfavorable attitudes toward the crippled.

Crippled child, for this study, is defined as a child of elementary school age who is visibly disabled, through congenital or acquired defects, in the use of this legs and body musculature so that he is unable to compete on terms of equality with a normal individual of the same age in the matter of locomotion.

Normal child refers only to the normality of not having an obvious crippling disability as referred to above.

Disability is used throughout this study as a static term, merely indicating that the child has some defect or impairment.

Handicap is a more dynamic term referring to the degree to which the defect or impairment interferes with, or limits, the individual's reaching his maximum. (A physical defect may cause a psychological handicap.)

Projective technique, "The essential feature of a projective technique is that it evokes from the subject what is, in various ways, expressive of his private world and personality process." (2:47) "Tell me a story...", and "Finish this sentence..." are variations of widely used projective techniques.

Generalize refers here specifically to the tendency of the noncrippled to "spread" evaluation of characteristics actually affected by the injury to other characteristics not so affected.

Adjustment rating is the score given by teachers to indicate the child's degree of social adjustment in the school situation.

Sample Used in This Study

Fifty-four elementary school children were chosen randomly as subjects for this study, eighteen from each of the three grade levels. These were evenly divided among three schools representative of three distinct "socio-economic" levels. Six children from each of the grades in each school comprised the sample.

Procedure

Two projective techniques, especially adapted for this study, were administered to each of the fifty-four subjects in an effort to identify existing attitudes (and to explore possible factors influencing their development) toward crippled children.

1. Instrument A, comprised of two parts, employed a "Tell me a story" technique; a picture of a seven-year-old girl was used as a stimulus. Children were asked to write stories about the girl in the picture. Some time later, a similar picture was shown and was presented in a similar manner. This time, however, the teacher casually referred to the subject in the picture as "a little crippled girl." Differences between the two stories comprised the data for assessing the attitudes.
2. Instrument B was a "Complete this sentence" test wherein three sentence "starters," among the ten in the test, were related to cripples.

Analysis of the data from these two instruments indicated that responses seemed to fall into two rather well-defined classifications:

1. Social responses, indicating acceptance or rejection of the crippled person such as: "I play with him;" "She is my friend;" "We won't let him play with us;" etc.
2. Value responses, indicating a judgment of the crippled person such as: "He is no good;" "He is bad;" "She can't do anything;" etc.

These general classifications were then formulated by the judges into two scales: 1) acceptance-Rejection Scale, 2) Worthy-Devalued Scale.

The first term in the name given each of the scales devised was the term indicating the "ideal" attitude toward cripples. The second term in the name given each scale thus indicated the opposite pole in the continuum. Four categories (degrees of intensity of attitude) were set up and coding procedures established. A four-point scale was used because of the widely recognized weakness of odd-numbered scales. The most desirable pole of each scale continuum was given the smallest score, one (very favorable attitude), and the least desirable was given the highest score, four (very unfavorable attitude). The large number of responses falling in the middle range were differentiated by being given a score of two (somewhat favorable) or three (somewhat unfavorable).

Each judge worked alone and evaluated the two stories independently. Stories were rated on two counts:

1. The general "feeling tone" of the story. This included an analysis of words and phrases instrumental in setting the "mood." For example, the general feeling tone might be one in which the reader was left with the general impression that the writer was neutral, or at least not unfavorable toward the central character in the story about the crippled girl. However, if the story should be rife with expressions such as "dark, dismal day," "frowning," "weeping," "lonely," etc., then the rating would tend to drop toward a "somewhat unfavorable" rating score.
2. The general "ending" of the story. This involved such questions as: "Is the cripple seen as able to succeed in life, though crippled?" "Does the future look hopeless for the cripple?" "Is the ending 'realistic and acceptable'?" "Does the subject provide a convenient 'miracle' ending so that he does not have to deal with a crippled heroine?"

Inasmuch as these two criteria were difficult to differentiate in some cases, it was decided to find the mean of the scores as a more consistent and dependable measure when the two scores did not agree. Thus, if a judge rated the "feeling tone" a "three" and the ending a "two," his mean evaluation of this attitude as revealed in the story would be a "two-point-five." The general rating in this case would be identical with the judge who rated the "feeling tone" a "two" and the "ending" as a "three." A general rating was given for both A-R and W-D scales.

The Sentence Completion Test was rated, also, on the four-point scales, on A-R and W-D scales. All three of the significant sentences were judged as a group. For example, if the sentences were

TABLE I

NUMBER OF A-R AND W-D* RATINGS ON INSTRUMENT A AND INSTRUMENT B BY GRADE LEVELS
(N = 18)

Instrument A	Grade 1 Score				Grade 3 Score				Grade 6 Score				Total			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Story 1																
A-R	0	18	0	0	0	16	2	0	0	17	0	1	0	51	2	1
W-D	0	17	1	0	0	17	1	0	0	15	2	1	0	49	4	1
Story 2																
A-R	1	10	7	0	0	2	15	1	0	1	16	1	0	14	38	2
W-D	1	11	6	0	0	2	15	1	0	2	16	0	1	15	37	1
Instrument B																
A-R	1	11	6	0	1	2	14	1	0	1	16	1	2	14	36	2
W-D	1	10	7	0	1	3	13	1	0	2	15	1	2	15	35	2

* A-R = Acceptance-Rejection Scale
W-D = Worthy-Devalued Scale

TABLE II

COMPARISON OF ATTITUDES ON THE ACCEPTANCE-REJECTION SCALE TO STORIES 1 AND 2 ON INSTRUCTION A

Story	Unfavorable Attitudes	Favorable Attitudes	Total
1	3	51	54
2	40	14	54
Total	43	65	108

χ^2 = 40.820 df = 1 P. = < .01

TABLE III

COMPARISON OF ATTITUDES ON THE WORTHY-DEVALUED SCALE TO STORIES 1 AND 2 ON INSTRUMENT A

Story	Unfavorable Attitudes	Favorable Attitudes	Total
1	5	49	54
2	38	16	54
Total	43	65	108

χ^2 = 40.401 df = 1 P. .01

FIGURE I

GRAPH SHOWING INCREASE IN "UNFAVORABLENESS" AND DECREASE OF "FAVORABLENESS" IN ATTITUDES TOWARD CRIPPLES BETWEEN GRADES 1, 3, AND 6

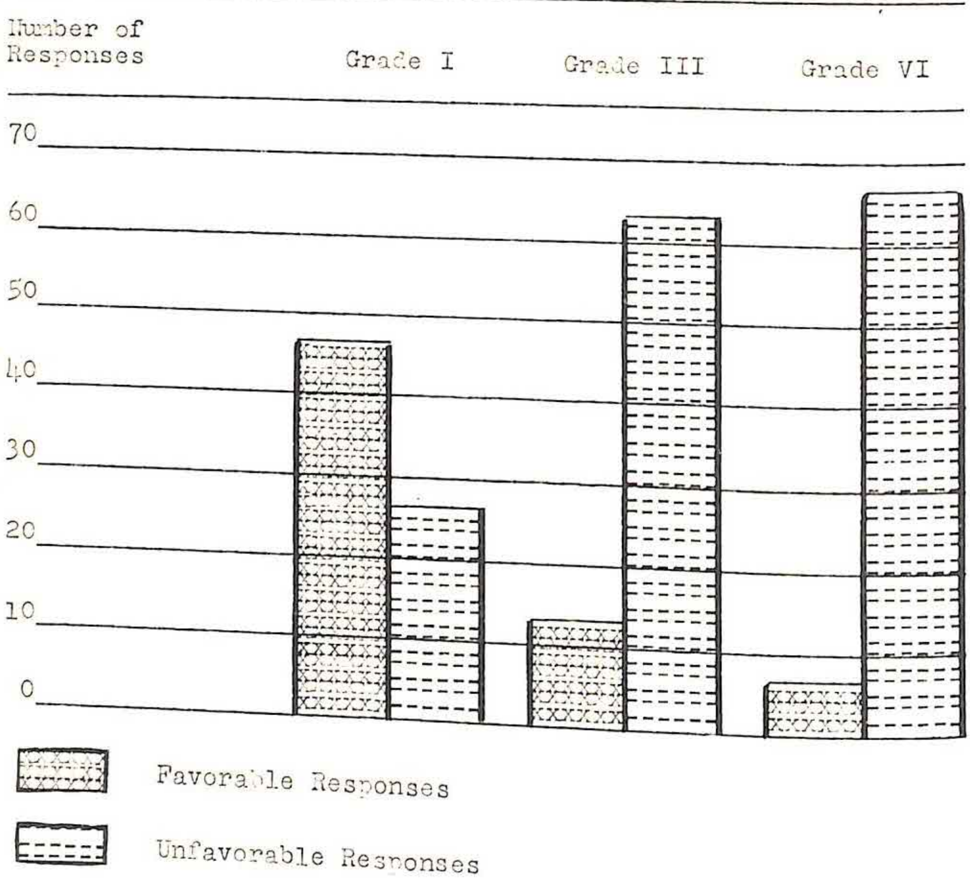


TABLE IV

COMPARISON OF ATTITUDES WITH ADJUSTMENT RATINGS ON ACCEPTANCE-REJECTION SCALE, STORY 2, INSTRUMENT A

Adjustment Ratings	Unfavorable Attitudes	Favorable Attitudes	Total
High	37	6	43
Low	6	5	11
Total	43	11	54

$\chi^2_c = 3.972$ df = 1 $P = < .05$

"Crippled kids are mean;" "The boy on crutches is helpless;" and "The girl in the wheelchair is crying," the words "mean," "helpless," and "crying" would provide the rater with the basis for judging the attitude.

Each teacher is required to evaluate a child's general social adjustment periodically throughout the school year. The most recent of these scores was used in testing Hypothesis III. It was an effort to discover if the child whom the teacher rated as "best" adjusted would be the "most favorable" in his attitudes toward cripples.

Analysis and Findings

The analysis of the data gave support to two of the three major hypotheses of this study, namely Hypothesis I and Hypothesis II.

> Hypothesis I—Attitudes of noncrippled children toward crippled children are significantly more unfavorable than their attitudes toward noncrippled children.

This hypothesis was tested through the use of a Chi Square Test (3:384). Data in Table I (below) indicate the number of scores in each category on both stories of Instrument A and on Instrument B. Tables II and III indicate the significance of differences between attitudes on Story 1 and Story 2 of Instrument A, which was used to test this hypothesis. Because of the small number of responses falling in categories 1 and 4 (the extremes), categories 1 and 2 were combined as indicating "favorable" responses and categories 3 and 4 were combined as "unfavorable" when computing the significance of the difference between attitudes on Story 1 and Story 2. This procedure was followed in all of the tables used in the testing of all hypotheses.

An examination of the data shown in Table I indicates that, in both the A-R and the W-D areas, by far the greater number of "favorable" ratings were found for Story 1 (noncrippled stimulus). An almost equally large number of "unfavorable" ratings were given to Story 2 (crippled stimulus). Conversely, a negligible number of "unfavorable" ratings were given to Story 1 and only a slightly greater number of "favorable" ratings were given to Story 2. (See Tables II and III.)

Thus Hypothesis I appeared to be substantiated from the data secured for this study. Elementary school children grades 1, 3, and 6 expressed attitudes toward crippled children that were significantly more "unfavorable" than their attitudes toward noncrippled children.

> Hypothesis II—Attitudes toward crippled children are a function of the grade level (age) of the child holding the attitudes.

A Chi Square Test (3:384) was applied to the data to test the above hypothesis. Comparisons between the various grades is summarized in Table I. The data revealed that there was a significant difference between grades 1 and 6. However, such a difference was not shown to exist between grades 3 and grades 6. The slight increase in the number of "unfavorable" responses between grades 3 and 6 indicated the same direction of change as between grades 1 and 3 but the difference was not statistically significant. Figure 1 depicts, in graphic form, the increase of "unfavorableness" and the decrease of "favorableness" as revealed by the data.

> Hypothesis III—Attitudes toward crippled children are a function of the social-emotional adjustment of the child holding the attitudes.

This hypothesis was tested by use of the same Chi Square Test (3:384). All responses on Story 2 of Instrument A and all responses to Instrument B, both A-R and W-D scales, were used. The same procedure was applied in the grouping of ratings as for Hypothesis I and Hypothesis II, namely categories 1 and 2 were combined to form the "favorable" attitudes and categories 3 and 4 were grouped to form the "unfavorable" attitudes. A similar procedure was used in classifying the adjustment ratings classroom teachers had made of the subjects. Those scores at or above the median (scores 3, 4, 5) were grouped to form the "high" adjustment ratings and those scores below the median (scores 1 and 2) were combined and considered as "low" adjustment.

Thus the variables for Hypothesis III were the adjustment ratings ("high" and "low") of classroom teachers and the scores on attitudes ("favorable" and "unfavorable") made by the judges in this study. Analysis of the data revealed that the hypothesis was not supported (see Tables IV, V, VI, and VII).

It had been predicted that those children judged to be "high" in adjustment would be found to be most "favorable" in their attitudes toward cripples. Rather than finding a positive relationship between these two variables, however, inspection of the data reveals that a negative relationship exists at the .01 level of confidence. That is, the students judged to be "high" in adjustment were the same students who were most "unfavorable" in their attitudes toward cripples.

Data presented in the tables indicate a markedly greater proportion of subjects rated "high" on adjustment held "unfavorable" attitudes toward cripples. Little difference was found between the "favorable" and the "unfavorable" attitudes of children who rated "low" on adjustment.

Summary of Findings

The data indicated that, in general:

1) Noncrippled children held unfavorable atti-

TABLE V

COMPARISON OF ATTITUDES WITH ADJUSTMENT RATINGS ON WORTHY-DEVALUED SCALE, STORY 2, INSTRUMENT A

Adjustment Ratings	Unfavorable Attitudes	Favorable Attitudes	Total
High	38	5	43
Low	5	6	11
Total	43	11	54

$\chi_c^2 = 7.402$ df = 1 P = < .01

TABLE VI

COMPARISON OF ATTITUDES WITH ADJUSTMENT RATINGS ON ACCEPTANCE-REJECTION SCALE, INSTRUMENT B

Adjustment Ratings	Unfavorable Attitudes	Favorable Attitudes	Total
High	38	4	42
Low	7	9	12
Total	45	9	54

$\chi_c^2 = 4.331$ df = 1 P = < .05

TABLE VII

COMPARISON OF ATTITUDES WITH ADJUSTMENT RATINGS ON WORTHY-DEVALUED SCALE, INSTRUMENT B

Adjustment Ratings	Unfavorable Attitudes	Favorable Attitudes	Total
High	38	5	43
Low	5	6	11
Total	43	11	54

$\chi_c^2 = 7.402$ df = 1 P = < .01

tudes toward their crippled peers.

2) The oldest group of children (age 11.8 years) were significantly more unfavorable in their attitudes toward cripples than the youngest group (age 6.5 years).

3) Children judged (by classroom teachers) to be "high" in adjustment were rated (by judges in this study) to be most "unfavorable" in their attitudes toward cripples.

The findings concerning the major hypotheses tended to point up the significance of attitudes displayed toward crippled children as a factor in the social-emotional adjustment of both able and disabled children.

Summary of Implications for the Theory

Barker's theory that "individuals who differ physically from the majority of people around them have the status of minority group members and, as such, are subject to the dynamics that hold for any minority group" (1:8), was corroborated by the findings in this study.

Status was made evident by the attitudes of others toward them and it was shown that the noncrippled children (majority group) do hold attitudes that are unfavorable toward crippled children (minority group) and those attitudes tend to reject and devalue the crippled child, forcing him into the position that Barker describes as "ambiguous" and "under-privileged", thus breeding in him insecurity, conflict and frustration. This study showed that there was a significant difference between attitudes which the elementary school children in this study displayed toward crippled children and those which they displayed toward noncrippled children.

Data revealed that crippled children were considered as "inferior" and that they were avoided or rejected by their peers. It further revealed that the older children (grades 3 and 6) were significantly more "unfavorable" in their attitudes than the younger children (grade 1). Hence, for children in grades 1, 3, and 6, the data supported Barker's theory that cripples have the status of minority group members. That they are also subject to the dynamics that hold for any minority group is also supported by this study.

When education in the public schools for all children is the common goal, when equal status is assumed in the classrooms of the elementary school, then the implication seems clear. Ways must be found to reduce prejudice toward cripples if they are not to be further handicapped by the unfavorable attitudes shown them by others in the school.

REFERENCES

1. Barker, R. C. "Somatopsychological Significance of Crippling," Bulletin 55 of the Social Science Research Council, New York, 1953, pp. 59-89.
2. Frank, L. K. Projective Methods (Springfield, Ill.: Charles Thomas Publications, 1948).
3. Edwards, A. L. Statistical Methods for Behavioral Sciences (New York: Rinehart and Co., 1954).
4. Meyerson, Lee. "Physical Disability as a Social-Psychological Problem," Journal of Social Issues, IV (1948), Editorial Notes.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

FACTORS ASSOCIATED WITH ABSENTEEISM AMONG STUDENTS IN TWO METROPOLITAN HIGH SCHOOLS

JAMES E. GREENE, SR.
The University of Georgia

The Problem

THE PRESENT STUDY deals with these two hypotheses:

1. Significant demographic differences exist between students having "best" attendance records versus students having "worst" attendance records in two metropolitan public high schools.
2. The factors associated with absenteeism vary significantly according to whether the school is located in an economically "advantaged" school community (School A) or in an economically "disadvantaged" school community (School D).

The Research Design

In both schools, weighted samples of "best" attenders and "worst" attenders were compared with respect to 1) scores earned on a specially selected battery of psychological tests, and 2) miscellaneous demographic data (i. e., age, school marks, etc.). The "worst" attenders consisted of the ten percent of students in each grade, by each sex separately, who had the highest number of absences during the semester preceding the present study. Conversely, "best" attenders consisted of the ten percent of students of each grade and sex who had fewest absences or no absences during the preceding semester. The proportion of individuals in each grade and of each sex constituting each of the two samples was the same as that of the given grade and sex in the total school population.

The basic data used in the comparisons of "best" versus "worst" attenders in School A and School D were as follows:

1. Problem Check List - (Mooney)
2. Greene's adaptation of Roger's Adjustment Inventory
3. Adjustment-Inventory (Bell)
4. School Adjustment Inventory - (Bell)
5. Absences from school
6. Chronological age
7. Intelligence quotient
8. Mark in English
9. Socio-economic status - (Sims)
10. Parental opinion of school - (Hand)

All of the test data (with the exception of intelligence quotient, which was secured from school records) were obtained from a special testing program devised and administered by the author. The remaining data were obtained from official school records.

All tests and instruments were administered according to official directions provided by the respective manuals. Test scores were double checked for accuracy. The basic data were edited, coded, and punched on IBM punch cards for processing by conventional punch card procedures. Inter-group comparisons of "best" attenders versus "worst" attenders in both schools were made for each sex separately. All obtained such differences were tested for statistical significance.

Summary of Findings

The findings are organized in a series of tables with reference to the two broad hypotheses previously listed. Because of space limitations, the interpretations of the findings will be limited to certain broad generalizations based on statistically significant differences between "best" and "worst" attenders and to a few specific detailed findings which appear to merit special comment.

Problem sensitivity (Mooney). In School A (economically "advantaged" community), none of the Mooney problem categories were significantly related to attendance among girls; however, best attending boys had significantly lower scores than did worst attending boys on (a) adjustment to school work, (b) total problems checked, and (c) finances, living conditions and employment (Table I).

TABLE I

COMPARISON OF SCORES ON MOONEY PROBLEM CHECK LIST EARNED BY "BEST" ATTENDERS VERSUS "WORST" ATTENDERS IN SCHOOLS "A" AND "D"; BY SEX

Mooney Problem Categories	School	Boys "Best" M	Boys "Best" SD	Boys "Worst" M	Boys "Worst" SD	Boys M_B-M_W	Boys "t"[a]	Girls "Best" M	Girls "Best" SD	Girls "Worst" M	Girls "Worst" SD	Girls M_B-M_W	Girls "t"[a]
Health and physical development	A	3.08	2.22	3.69	2.96	-.61	1.27	3.28	2.20	3.77	2.12	-.49	.70
	D	2.25	2.29	2.66	2.59	-.41	.80	3.17	2.45	4.46	2.87	-1.29	2.61
Finances, living conditions, employment	A	3.15	2.36	4.19	3.29	-1.04	2.00	3.14	2.95	3.08	2.68	.06	.10
	D	3.05	2.60	4.41	2.83	-1.35	2.45	3.24	2.81	4.94	3.13	-1.70	3.04
Social and recreational activities	A	4.61	3.19	5.07	2.95	-.46	.82	4.82	3.24	4.77	3.28	.05	.08
	D	3.82	3.16	4.34	3.34	-.52	.79	4.48	2.94	4.44	3.28	.04	.07
Courtship, sex, marriage	A	3.54	2.87	4.10	2.93	-.56	1.06	3.64	3.01	4.41	2.90	-.77	1.31
	D	3.20	3.25	3.00	2.79	.20	.33	4.50	3.03	4.54	3.23	-.04	.05
Social-psychological relations	A	3.30	3.07	3.85	3.13	-.55	.98	4.74	3.22	4.73	2.95	.01	.02
	D	2.51	3.24	3.00	3.69	-.49	.69	4.08	3.52	4.38	3.34	-.30	.48
Personal-psychological relations	A	3.87	2.76	4.86	3.29	-.99	1.80	5.40	3.64	6.31	3.10	-.91	1.36
	D	2.31	2.79	2.43	2.63	-.12	.22	3.82	2.58	4.62	3.35	-.80	1.43
Morals and religion	A	3.74	2.66	4.15	2.94	-.41	.84	3.70	3.33	4.10	2.76	-.40	.67
	D	2.13	2.91	2.00	2.23	.13	.25	2.31	2.56	3.42	3.09	-1.10	2.08
Home and family	A	2.57	2.66	3.32	2.85	-.75	1.50	3.16	3.45	3.37	3.04	-.21	.32
	D	3.18	3.01	3.27	3.08	-.09	.15	2.97	2.93	3.25	3.07	-.28	.50
Future: vocational and educational	A	3.74	3.02	4.73	3.40	-.99	1.52	3.80	3.16	3.79	3.05	.01	.02
	D	4.76	3.40	4.82	2.74	-.05	.08	5.08	2.93	5.98	3.20	-.90	1.58
Adjustment to school work	A	5.77	3.13	7.32	3.20	-1.55	2.98	5.54	3.58	6.33	3.55	-.79	1.13
	D	3.51	3.22	3.09	2.79	.42	.69	2.73	2.63	2.85	2.44	-.12	.33
Curriculum and teaching procedures	A	5.70	.43	6.24	.46	-.54	.86	7.66	3.03	7.62	3.03	.04	.07
	D	5.37	3.67	5.20	3.27	.17	.24	5.21	3.47	4.81	3.77	.33	.49
Total problems checked	A	30.25	23.60	56.02	27.66	-25.77	5.48	51.00	25.90	57.50	25.10	-6.50	1.29
	D	36.82	24.90	39.09	24.90	-2.27	.45	44.40	26.50	50.00	28.90	5.60	1.08
Total problems circled	A	10.25	11.83	11.11	17.11	-.86	.32	13.64	13.82	14.77	15.17	-1.13	.39
	D	7.18	13.04	8.64	13.45	-1.46	.54	14.24	18.70	11.54	14.90	2.70	1.87

[a] A "t" at 1.98 and above is significant at .05 level.

In School D (economically "disadvantaged" community), worst attending boys had higher scores on finances, living conditions and employment than did best attending boys. Best attending girls had significantly lower scores than did worst attending girls on the following Mooney categories: (a) health and physical development, (b) finances, living conditions and employment, and (c) morals and religion (Table I).

Self-assessed status and aspiration (Rogers - Greene). Among boys enrolled in School A, good attendance was significantly related to favorable self-ratings of status on "learns quickly". Among girls enrolled in School A, good attendance was significantly related to favorable self-ratings on (a) being well liked by teachers, (b) learns quickly, (c) aspires to learn quickly, and (d) total status score (Table II).

Among boys in School D, none of the self-assessed measures of status or of aspiration was significantly related to attendance. Among girls in School D, good attendance was significantly related to favorable self-ratings on (a) total status score and (b) total aspiration score (Table II).

Standardized measures of adjustment (Bell). The data in Table III are based on sub-scores and total score on the Bell Adjustment Inventory and total score on the Bell School Adjustment Inventory.

Among boys enrolled in School A, none of the categories of adjustment shown in Table III were significantly related to attendance. Among girls enrolled in School A, good attendance was associated with favorable scores on the following categories of Bell's Adjustment Inventory: (a) health adjustment, (b) emotional adjustment, (c) general adjustment (i. e., total score).

Among boys enrolled in School D, none of the categories of adjustment was significantly related to attendance. Among girls enrolled in School D, good attendance was associated with the following categories on the Bell Adjustment Inventory: (a) home adjustment, (b) health adjustment, (c) general adjustment.

Miscellaneous variables. Among both boys and girls enrolled in School A, good attendance was associated positively with high marks in English. (Table IV).

Among both boys and girls in School D, good attendance was associated with (a) comparatively high intelligence quotients, (b) comparatively high marks in English, (c) comparatively high socio-economic status (Sims), and (d) comparatively favorable parental opinions of the school (Hand).

The gross amount of non-attendance during the semester on the part of the "worst attending" groups among both sexes of both schools was of such magnitude as presumably to interfere seriously with the effectiveness of school work. When expressed as a percent of the total number of school days during the semester, the mean absences of the "worst attending" groups ranged from 16.3 percent (School D, girls) to 21.1 percent (School D, boys).

Inter-school differences in factors associated with absenteeism. The "t" values of the inter-attendance group differences shown in Tables I-IV provide a comparison of the extent to which the factors which reliably differentiate between "best" and "worst" attenders of each sex vary from one type of school community to the other. These data appear to warrant the following broad generalizations:

1. Among both sexes of both schools, "best" attenders earn reliably higher school marks than do "worst" attenders. Only English marks are shown in Table IV. Supplementary data (not shown in Table IV) indicate that favorable marks in other in all school subjects were reliably associated with good attendance.

2. Among students of both sexes in School D (i.e., semi-slum zone), good attendance is reliably associated with high I. Q., favorable socio-economic status and favorable parental opinion of the school. The comparable findings for School A (i. e., economically advantaged community) are either inconsistent or unreliable. Favorable self-estimated learning capacity is reliably associated with good attendance in School A but not in School D. When it is recognized that the higher the numerical score the more favorable the status on the Sims socio-economic scale and on the Hand inventory of parental opinion of school (on which negative scores represent "unfavorable" opinions) it is seen that School D community has a markedly "disadvantaged" status in these two respects as compared with School A community. Under these circumstances, it is not surprising that among students in a semi-slum zone (School D), a comparatively low intelligence quotient, a comparatively low socio-economic status of the home, and a comparatively unfavorable parental opinion of the school would be more closely and reliably associated with absenteeism than among students living in an "economically advantaged" community (School A).

3. Among girls in both schools, good attendance is reliably associated with favorable scores on health adjustment, emotional adjustment, and general adjustment, whereas the comparable findings for boys are inconsistent or unreliable.

4. The "t" values for the remaining inter-school comparisons are either too inconclusive or too inconsistent to warrant generalization.

Conclusions and Implications

In general, the detailed findings support the conclusion that absenteeism is a symptomatic behavior associated with certain other independent variables, each of which tends to be individually

TABLE II

COMPARISON OF SELF-ASSESSED STATUS AND ASPIRATIONS OF "BEST" ATTENDERS VERSUS "WORST" ATTENDERS IN SCHOOLS "A" AND "D"; BY SEX

Self-assessed[a] status and aspiration categories	School	Boys "Best" M	Boys "Best" SD	Boys "Worst" M	Boys "Worst" SD	M_B-M_W	"t"[b]	Girls "Best" M	Girls "Best" SD	Girls "Worst" M	Girls "Worst" SD	M_B-M_W	"t"[b]
Examinee a leader ?													
Status	A	3.00	1.79	2.65	1.71	.35	1.06	2.56	1.44	2.14	1.33	.42	1.50
	D	3.10	2.06	3.59	1.76	-.49	1.26	2.47	1.71	2.44	1.84	.03	.11
Aspiration	A	4.37	2.25	4.39	2.23	-.02	.05	3.27	1.92	3.13	1.88	.14	.37
	D	4.42	2.19	4.67	1.83	-.25	1.61	3.76	2.41	3.52	2.26	.24	.55
Examinee popular with other students ?													
Status	A	3.89	1.87	4.09	2.18	-.20	.53	3.98	1.70	3.80	1.93	.18	.47
	D	4.38	2.09	4.11	1.81	.27	.69	3.48	2.03	3.56	2.47	-.08	.19
Aspiration	A	5.90	1.49	5.59	1.70	.31	1.07	5.66	1.12	5.37	1.91	.29	.94
	D	5.69	1.72	5.26	1.87	.43	1.16	6.26	1.11	6.08	1.38	.18	.75
Examinee well liked by teachers ?													
Status	A	4.05	1.97	3.67	2.22	.38	.97	4.72	1.88	3.67	2.04	1.05	2.63
	D	4.42	2.36	4.27	1.78	.15	.50	5.14	2.04	4.36	2.37	.78	1.86
Aspiration	A	4.73	1.91	4.81	2.05	-.08	.22	5.40	1.51	4.90	2.05	.50	1.32
	D	5.50	1.84	4.95	1.88	.55	1.45	5.92	1.66	5.82	1.67	.10	.31
Examinee learns quickly ?													
Status	A	3.60	1.98	2.86	1.91	.74	2.06	4.06	1.98	3.00	1.93	1.06	2.65
	D	3.87	2.18	4.33	1.80	-.46	1.12	4.20	2.28	3.58	2.30	.62	1.68
Aspiration	A	6.08	1.26	6.14	1.29	-.06	.25	5.87	1.28	6.29	.69	-.42	2.00
	D	5.85	1.66	6.09	1.13	-.24	.83	6.20	1.21	5.86	1.72	.34	1.21
Examinee well treated by his family ?													
Status	A	6.00	1.33	6.05	1.54	-.05	.21	5.84	1.35	6.17	1.09	-.33	1.38
	D	5.96	1.41	5.95	1.33	.01	1.04	6.09	1.42	5.61	1.61	.48	1.81
Aspiration	A	6.31	1.14	6.29	1.12	.02	.09	6.19	.98	6.38	1.00	-.19	.90
	D	5.75	1.79	6.07	1.30	-.32	1.00	6.16	1.47	5.88	1.74	.28	.87
Total adjustment	A	20.99	6.06	20.86	5.61	.13	.12	21.13	1.65	18.92	1.79	2.21	2.08
	D	21.05	6.39	22.79	4.47	-1.74	1.55	22.06	5.70	19.65	6.48	2.41	2.08
Total aspiration	A	27.65	5.68	27.27	5.75	.38	.35	26.64	1.18	26.14	1.78	.50	.51
	D	27.29	6.36	27.64	4.80	-.35	.32	28.73	5.19	25.22	5.76	3.51	3.38
Adjustment-aspiration discrepancy	A	-9.27	6.06	-8.70	8.12	.57	.42	-7.26	1.28	-9.41	1.01	-2.15	1.70
	D	-7.50	7.10	-6.31	6.45	1.19	.85	-8.52	7.30	-7.81	9.95	.71	.42

[a] Scores on each of these categories are based on a 7-point self-rating scale (i.e., high score, 7; low score, 1).
[b] A "t" at 1.98 and above is significant at .05 level.

sympotomatic of an unfavorable adjustment between the learner and the educational and social environment in which he is operating.

It is perhaps appropriate to state that the author has had extensive experience in the clinical study of individual absentees in both School A and School D and in many other schools. These clinical studies support the conclusions, which are to some extent but not fully corroborated by the statistical findings herein reported, that the typical absentee is an individual exposed to conditions and pressures which produce unfavorable attitudes toward the school. These conditions are likely to include one or more of the following: unfavorable parental opinion of the school, a previous history of absences, low previous and current school marks, over-ageness in grade, dullness , arbitrary and inappropriate curricula, inflexible academic standards, and teachers who lack a sympathetic understanding of the adjustment problems confronting the individual offender and/or who are unskilled in creating in their classrooms an intellectual, emotional and social climate conducive to effective learning and adjustment.

As is true in all therapeutic programs, remedial action concerning absenteeism should be directed toward removing or alleviating the unique condition or conditions which produced the symptomatic behavior of the particular offender. The extensive prevalence of absenteeism and the conclusions concerning the correlates of this symptomatic behavior would appear to have important implications for the pre-service and in-service training programs of teachers, counselors, and administrators.

TABLE III

COMPARISON OF CERTAIN ADJUSTMENT SCORES EARNED BY "BEST" ATTENDERS VERSUS "WORST" ATTENDERS IN SCHOOLS "A" AND "D"; BY SEX

Adjustment Categories		Boys						Girls					
		"Best"		"Worst"				"Best"		"Worst"			
	School	M	SD	M	SD	$M_B - M_W$	"t"[a]	M	SD	M	SD	$M_B - M_W$	"t"[a]
Home	A	2.17	1.01	2.22	1.00	-.05	.33	1.80	1.52	1.91	.97	-.11	.65
	D	2.29	1.78	2.20	1.20	.09	.30	1.67	1.41	2.27	1.26	-.60	2.40
Health	A	1.84	.90	2.14	.80	-.30	1.88	1.62	.77	2.42	.73	-.80	5.33
	D	1.75	1.13	1.95	1.03	-.20	1.25	1.83	1.05	2.60	1.04	-.77	4.05
Social	A	1.84	.92	1.81	.88	.03	.19	1.76	.81	1.79	.82	-.03	.19
	D	1.80	.70	1.95	1.29	-.15	.16	2.11	.86	2.19	.81	-.08	.53
Emotional	A	2.06	1.08	2.21	1.09	-.15	.75	1.96	1.09	2.58	1.40	-.62	2.58
	D	2.05	1.18	1.86	1.12	.19	.82	2.17	1.04	2.56	1.09	-.39	1.95
General	A	2.11	.93	2.36	1.04	-.25	1.39	1.82	.93	2.33	.96	-.51	2.83
	D	2.22	1.02	2.25	.96	-.03	.05	2.17	1.04	2.67	1.00	-.50	2.63
School	A	3.07	.87	3.14	.91	-.07	.44	3.40	.80	3.63	.64	-.23	1.64
	D	2.95	.86	3.14	.95	-.19	1.06	2.73	1.07	2.69	1.20	.04	.19

Note: Low scores indicate favorable adjustment; high scores indicate unfavorable adjustment. The first five variables are based on the Bell's Adjustment Inventory. The school adjustment scores are based on the Bell's School Inventory.

[a] A "t" at 1.98 and above is significant at 105 level.

TABLE IV

COMPARISON OF "BEST" ATTENDERS VERSUS "WORST" ATTENDERS IN SCHOOLS "A" AND "D" ON CERTAIN MISCELLANEOUS VARIABLES; BY SEX

		Boys						Girls					
		"Best"		"Worst"				"Best"		"Worst"			
	School	M	SD	M	SD	M_B-M_W	"t"[a]	M	SD	M	SD	M_B-M_W	"t"[a]
Absences from school	A	1.00	2.09	14.92	6.60	-13.92	14.65	.72	1.70	15.85	7.56	-15.13	14.00
	D	1.60	2.53	20.55	9.10	-18.95	13.35	.38	3.85	14.73	7.30	-14.35	12.81
Chronological age	A	15.88	1.15	16.16	1.43	- .28	1.04	15.98	1.37	15.99	1.14	- .01	.04
	D	15.75	1.18	16.45	1.49	- .70	2.50	15.50	1.01	15.71	1.26	- .21	1.00
Intelligence quotient	A	106.42	15.50	103.65	13.60	2.77	.86	107.27	10.70	108.00	10.80	- .73	.32
	D	100.80	12.00	94.50	13.08	6.30	2.35	100.94	12.30	92.55	11.90	8.39	3.42
Marks[b] in English	A	3.25	1.24	2.53	1.08	.72	3.43	3.98	.99	3.35	1.13	.63	3.00
	D	3.29	1.00	2.10	.93	1.19	6.25	3.78	1.05	2.87	1.07	.91	4.79
Socio-economic status (Sims)	A	23.27	5.35	21.50	4.60	1.77	1.95	24.38	4.92	23.54	5.04	.84	.85
	D	15.94	4.64	13.31	4.56	2.63	2.80	16.05	6.16	12.65	5.28	3.40	3.21
Parental opinion[c] of school	A	26.05	16.00	20.83	19.50	5.22	1.12	20.69	20.20	14.76	21.10	5.93	1.00
	D	- 1.64	31.60	-32.27	24.90	-30.63	5.40	6.97	33.70	- 5.19	31.30	12.16	2.03

Note: School "A" is located in an economically "advantaged" school community, whereas School "B" is located in an economically "disadvantaged" school community. Both schools are in Atlanta, Georgia.

[a] A "t" at 1.98 and above is significant at .05 level.

[b] Based on arbitrarily weighted scores, A-5; B-4; C-3; D-2; E-1.

[c] Weighted scores derived from items in Illinois Inventory of Parent Opinion. Positive scores indicate a favorable parental opinion of the school; negative scores indicate an unfavorable opinion.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

THE KNOWLEDGE ABOUT PSYCHOLOGY TEST[1]

FRANK COSTIN[2]
The University of Illinois

INSTRUCTORS IN beginning psychology courses generally agree that one of their objectives is to have students develop an understanding of how the psychologist functions as a scientist and as a professional. This kind of knowledge necessarily involves not only the learning of specific facts, but the acquisition of more objective ways of looking at human behavior. It is reasonable to expect, therefore, that instructors might well be concerned with another kind of problem, one closely related to the development of objectivity – the problem of how students perceive psychologists as persons. For example, do students with little or no experience in psychology courses believe that psychologists are distinctly "different" in their "personalities," as a group, from people in other professions and business occupations? If this belief is consistent, it probably represents a stereotyped way of looking at behavior. A relevant goal of an introductory psychology course, therefore, would be to reduce this kind of perception. The purpose of this report is to describe the development of an instrument which appears to be useful for evaluating the attainment of such a goal. It is called the Knowledge About Psychology (KAP) Test, and is designed to measure (a) students' perceptions of psychologists' personal characteristics, and (b) how these perceptions are related to the more conventional kinds of knowledge students are expected to acquire concerning the scientific and professional characteristics of psychology.

Description of the Knowledge About Psychology (KAP) Test

In its present form, the KAP Test consists of two parts: Scientific and Professional Characteristics of Psychology (Scale S) and Personal Characteristics of Psychologists (Scale P).

Scale S (30 items) is intended to measure the student's understanding of the scientific nature of psychology, and some of its characteristics as a profession today, with emphasis on the special functions carried out by the psychologist in his role as a scientist and a professional person. The items in this scale are quite similar to those found in many classroom tests of achievement in beginning psychology courses. The S Scale, however, serves a second purpose – as a disguise for Scale P.

Scale P (30 items) is designed to measure what students think psychologists are like "as people", independent of their professional and scientific roles. The chief aim in constructing items for this scale was to provide a way of evaluating the extent to which students accepted stereotyped beliefs or misconceptions about the "personalities" of psychologists. Such beliefs involve the basic stereotype that in their personality traits and private lives, psychologists are distinctly different, as a group, from other professional workers, from businessmen, and in certain instances, from the general population. The underlying rationale for this scale follows a conclusion reached by Anne Roe (2:230) concerning the "personalities" of scientists:

"Scientists are people, not rational automatons. They differ from other people in terms of what they do, in the things that give them satisfaction, more than in terms of completely special capacities. There is nothing you can say about them as persons that you cannot also say about some people who are not scientists."

The following sample items illustrate the general nature of the two scales:

Examples of items from Scale S:

1. Which of the following usually contribute most to scientific research?
   A. psychologists B. psychiatrists
   C. psychoanalysts

2. To be scientific, a psychological experiment must:
   A. discover important facts B. control factors other than those being investigated
   C. take place in a laboratory

3. Psychologists sometimes use animals for experiments because:
   A. animal behavior is less complex than human

behavior
B. facts obtained from observing animals are usually true of human behavior
C. studying animal behavior is more scientific than studying human behavior

4. The psychologist is a scientist because he:
A. makes systematic measurements of human behavior
B. uses his discoveries to help people in their everyday problems
C. speculates about the human mind

Examples of items from Scale P:

1. As compared with business men in general, psychologists are on the average:
A. less poised B. more precise
C. about as poised and precise

2. Three men are dining together at a restaurant. Man No. 1 is very interested in the food. Man No. 2 spends a great deal of the time admiring the furnishings. Man No. 3 pays most attention to the other people in the restaurant. The three men are a psychologist, a corporation executive, and an architect. In identifying them which of the following statements is most likely to be true?
A. No. 1 is the corporation executive, No. 2 is the architect, and No. 3 is the psychologist
B. No. 3 is the psychologist, but it is difficult to distinguish between the corporation executive and the architect.
C. No. 3 could very well be any one of the three men

3. As a group, psychologists are:
A. less self-confident than businessmen and lawyers
B. about as self-confident as businessmen and lawyers
C. less self-confident than lawyers

4. Children of psychologists generally enjoy spending time with their fathers because:
A. most children enjoy spending time with their fathers
B. psychologists are able to tell their children interesting stories
C. psychologists are able to help their children understand themselves

Development of the Test Items

The chief source of ideas for the general format of the KAP Test, and for the specific items, particularly with respect to Scale P, was the Facts About Science (FAS) Test (5)b. With appropriate modifications, many of the FAS Test items, which deal mostly with biological and physical science, and with personal characteristics of "scientists" in general, were very suitable for the purposes of the KAP Test. Other ideas for items came from data contained in studies by Bendig and Hountras (1), Terman (5), and Roe (2, 3); and from standard introductory textbooks in psychology. Further information helpful in constructing items was obtained by having students enrolled in an introductory course reply to questionnaires administered at the beginning of the course. These questionnaires requested the students to describe what they thought a psychologist was and did, how his work was different from that of other scientists, and how a psychologist was personally the same as or different from other professional workers and from people in business and industry.

On the basis of item analyses as well as judgments made by a panel of psychologists, a pool of over 100 items was reduced to 60 items, 30 for each scale. No item was retained for the final form unless all 10 members of the panel agreed with the author's keyed answer. A simplified procedure described by Stanley (4, pp. 436-53) was used to determine difficulty levels and discrimination values of the items. Table I shows this information for the present form of the test.

In obtaining subjects for the item analyses, intact classes enrolled in a wide variety of courses were tested. While all of these courses were offered in the College of Liberal Arts and Sciences, they were open to students from other colleges as well. The total sample thus obtained contained students from seven different colleges on the University of Illinois campus. All of the students tested were then classified according to their college class level and the number of psychology courses they had completed at the time of taking the KAP Test. Table I shows the results of item analyses for nine of these groups. Originally, the number of students in each of these 9 groups ranged from 73 for the smallest to 97 for the largest. For convenience of computation, and to make comparisons among groups more meaningful, each was reduced to an N of 70 by means of a random selection of cases from each group.

The data in Table I indicate that insofar as criteria of discrimination and difficulty of items are concerned, the present form of the KAP Test is reasonably satisfactory as a research instrument for evaluating the specified outcomes of the first course in psychology. Scale P seems to be a more discriminating and more difficult set of items than Scale S.

It will be noted that this is a rather short test. It was deliberately made brief so as to permit all students to complete it within an ordinary class period. This was done because it was anticipated that most instructors who would want to use this instrument would probably wish to do so during regular class time. In a sample of over 1000 students, no one required more than 40 minutes

TABLE I

DIFFICULTY LEVEL AND DISCRIMINATION VALUE OF ITEMS ON THE KNOWLEDGE ABOUT PSYCHOLOGY TEST, ACCORDING TO COLLEGE CLASS AND NUMBER OF PSYCHOLOGY COURSES COMPLETED

College Class	Number of Psychology Courses Completed N = 70 ea. group	Number of Items: Scale S* Easy	Scale S* Moderately Difficult	Scale S* Very Difficult	Scale S* D	Scale P** Easy	Scale P** Moderately Difficult	Scale P** Very Difficult	Scale P** D
Freshmen	0	5	19	6	15	3	21	6	22
Sophomore	0	5	21	4	17	6	23	1	17
	1	7	20	3	10	11	18	1	16
Junior	0	3	21	6	9	2	23	5	18
	1	11	16	3	8	9	19	2	9
	2 or more	13	15	2	9	13	16	1	8
Senior	0	6	18	6	9	5	24	1	15
	1	13	15	2	13	11	18	1	13
	2 or more	14	15	1	5	13	17	0	12

Note: Each of the two scales contains 30 items. Difficulty levels are according to the percentage of individuals missing a given item: "Easy", < 17%; "Moderately Difficult", 17% to 83%; "Very Difficult", > 83%. "D" indicates highly discriminating items. ($p < .05$).

*Scientific and Professional Characteristics of Psychology.

**Personal Characteristics of Psychologists.

TABLE II

STUDENTS' ACHIEVEMENT ON THE KNOWLEDGE ABOUT PSYCHOLOGY TEST ACCORDING TO THEIR THEIR COLLEGE CLASS AND NUMBER OF PSYCHOLOGY COURSES COMPLETED (N = 70 FOR EACH MEAN SCORE)

College Class	0: Mean Score	0: S. D.	1: Mean Score	1: S. D.	2 or more: Mean Score	2 or more: S. D.
Freshman						
Scale S	18.70	3.22	21.27	3.50	-----	----
Scale P	20.17	4.51	22.94	3.57	-----	----
r of scores	.45		.47			
Sophomore						
Scale S	20.37	2.82	22.61	2.81	-----	---
Scale P	21.06	4.70	23.70	3.43	-----	---
r of scores	.36		.29			
Junior						
Scale S	19.84	3.21	23.31	2.56	23.97	2.35
Scale P	20.21	4.45	23.83	2.73	25.11	2.79
r of scores	.40		.34		.39	
Senior						
Scale S	20.71	2.70	23.10	3.44	24.40	2.19
Scale P	21.93	4.07	23.93	3.80	24.70	3.08
r of scores	.48		.49		.26	

Column group heading: Number of Psychology Courses Completed by Students in Each Class

Note:
Scale S = Scientific and Professional Characteristics of Psychology (30 items)
Scale P = Personal Characteristics of Psychologists (30 items)
An individual's score on a scale is the number of items answered correctly.

TABLE III

DIFFERENCES IN MEAN SCORES ON THE KNOWLEDGE ABOUT PSYCHOLOGY TEST ACCORDING TO COLLEGE CLASS AND NUMBER OF PSYCHOLOGY COURSES COMPLETED

Differences within each class according to number of courses completed

Scale S. Scientific and Professional Characteristics

	Fr. -1	So-1	Jr. -1	Jr. -2+	Sr. -1	Sr. -2+
Fr. 0	2.57**					
So- 0		2.24**				
Jr. -0			3.47**	4.13**		
Jr. -1				0.66*		
Sr. -0					2.39**	3.69**
Sr. -1						1.30**

Scale P. Personal Characteristics

	Fr. -1	So-1	Jr. -1	Jr. -2+	Sr. -1	Sr. -2+
Fr. -0	2.77**					
So-0		2.64**				
Jr. -0			3.61**	4.90**		
Jr. -1				1.28**		
Sr. -0					2.00**	2.77**
Sr. -1						0.77

Differences between classes having completed the same number of courses

Scale S. Scientific and Professional Characteristics

	So-0	So-1	Jr. -1	Jr. -1	Sr. -0	Sr. -1	Sr. -2
Fr. -0	1.67**		1.14*		2.01**		
Fr. -1		1.34**		2.04**		1.83**	
So-0			-.53		.34		
So-1				.70		.49	
Jr. -0					.87		
Jr. -1						-.21	
Jr. -2+							.43

Scale P. Personal Characteristics

	So-0	So-1	Jr. -0	Jr. -1	Sr. -0	Sr. -1	Sr. -2+
Fr. -0	.89		.04		1.76*		
Fr. -1		.76		.89		.99	
So-0			-.85		.87		
So-1				.13		.23	
Jr. -0					1.72*		
Jr. -1						-.10	
Jr. -2+							-.41

Note: The digit accompanying each college class abbreviation indicates the number of psychology courses completed by that group. A positive difference favors the group in the row, a negative difference the group in the column. N = 70 for each group. See Table II for the mean scores upon which differences are based.

* $1.98 < t < 2.62$, $.01 < p < .05$

** $t > 2.62$, $p < .01$

to complete the test. The average time taken by these students was about 30 minutes. No time limit however, was stipulated in administering the test.

Reliability

Retest reliability (six weeks interval) was measured for 51 students enrolled in English and physical science courses during a summer session. Those students represented all four college class levels, and had never taken courses in psychology, nor were they taking any at the time. The r's obtained were .70 for Scale S and .78 for Scale P. A second estimate of retest reliability, also for a six weeks interval, was calculated for another group of 77 students, also enrolled in English and physical science. All of them had previously completed at least one course in psychology, but were not taking any at the time the KAP Test was administered. All four college class levels were represented in this group. The r's were .75 for Scale S and .81 for Scale P.

Using the Kuder-Richardson Formula 20, a coefficient of internal consistency was computed for the responses of 280 students, 70 from each of the four college class levels. None of these students had ever taken a course in psychology. For Scale S the r was .45, and for Scale P it was .76.

The reliability data seem to be consistent with the discrimination and difficulty level findings already discussed, in that Scale P seems to be a somewhat more successful instrument than Scale S. The internal consistency coefficients suggest that Scale P may be measuring a more homogeneous cluster of traits than Scale S.

Validity

A satisfactory degree of content validity for the KAP Test can be inferred on the basis of the sources of material used in developing the test, the judgments of the various psychologists used in the revision or rejection of items, and the item analysis data previously discussed (Table I). The prerequisite of reliability also seems to have been established to a reasonable extent, particularly for Scale P.

Some evidence also supports an inference of construct validity in the increase in mean scores as a function of the number of psychology courses taken, as shown for the same student groups used for the item analysis previously described and an additional group of 70 freshmen who had completed one course in psychology. Table II shows the mean scores and standard deviations obtained for these ten groups of students, and Table III shows the differences among these means.

These results indicate that students who had completed courses in psychology attained significantly higher mean scores on the S Scale than did those who had not taken psychology. These differences are shown in the upper half of Table III "(Differences within each class according to number of courses completed"). All eight of the differences were significant. Furthermore, they occurred not only when students with no courses in psychology were compared with those having completed one course, but also when students with one course were compared with those having completed two or more. The two differences of the latter kind, however, which involved comparisons of juniors (1 vs 2 or more courses) and comparisons of seniors (1 vs 2 or more courses) were the smallest of the differences obtained.

Very similar results were obtained for Scale P. The one exception was the non-significant difference between seniors with one course and seniors with two or more courses completed.

The above findings for differences among mean scores on the KAP Test according to number of psychology courses completed were consistent with the results previously presented concerning the item analysis of the test. (Table I). It will be recalled that the difficulty levels of the items were consistently higher for students who had not taken psychology than they were for students who had completed one or more courses.

Experience in psychology, then, seems to contribute significantly to higher test scores on the KAP Test. Before this conclusion can be accepted, however, another problem must be considered – the relationship between formal educational development as a whole and achievement on the KAP Test. That is, it may be that a comparison of students representing various degrees of educational development, as defined by formal college class level, would show that test scores improved with educational advancement just as much as they did with increases in the number of psychology courses completed. The lower half of Table III ("Differences between classes having completed the same number of courses") shows the results of examining this possibility. In the case of Scale S, only those comparisons involving freshmen revealed differences in mean scores according to college level. The freshmen groups scored significantly lower than did students in the other class levels with the same number of completed psychology courses. (It should be noted, however, that these differences were consistently smaller than those involving comparisons of means according to number of courses completed.) In the remaining seven comparisons among students of different educational levels, no significant differences in means occurred.

When the same kinds of comparisons were made for Scale P (Table III, lower half) only two out of 13 differences in mean scores were significant. Both of them involved groups with no psychology courses completed, seniors scoring higher than

freshmen and juniors.

All in all, then, experience in psychology was more relevant to achievement on the KAP Test than was general educational experience, as defined by college class membership. This finding was particularly true for achievement on Scale P.

Another empirical support for inferring construct validity (and content validity as well) can be shown in the correlations obtained between the two scales (Table II). Considering the total number of groups compared, the correlations in general are sufficiently low to warrant the conclusion that Scales S and P are measuring somewhat different kinds of knowledge and psychological processes—a guiding aim, on a logical basis, for the initial selection and construction of items.

Summary and Conclusions

The Knowledge About Psychology Test, whose development has been described in this article, together with evidence for its reliability and validity, appears to be a promising instrument for evaluating some important goals of psychology courses. In its present state it seems to be most suitable for use with underclassmen who have had little or no experience in psychology, and thus should have its greatest use in beginning courses.

The P Scale in particular seems to be especially worthwhile as a technique for assessing stereotyped perceptions and beliefs concerning psychologists as persons. While Scale S is apparently a measure of relatively straightforward factual information, Scale P is probably tapping, though in an indirect manner, the manifestations of attitudes closely related to stereotyped ways of thinking about human behavior.

The present investigator is now completing a study in which the KAP Test was used to evaluate changes which may occur as the result of a beginning course in psychology offered by the Division of General Studies, a two year general education program at the University of Illinois. Precourse and postcourse scores on the KAP Test will be analyzed to see how they are related to achievement on regular course examinations, to verbal aptitudes, to sex differences, and to other variables.[3] It is hoped that instructors interested in this kind of study might also wish to use the KAP Test to carry out similar evaluations. While some instructors might hypothesize with confidence that students completing an introductory psychology course would show appreciably higher scores on the S scale, since this kind of content is usually taught rather directly, they might have considerably less confidence in a prediction that significant increases would occur on the P Scale as well. Yet, while the kind of knowledge required to effect positive changes on this scale is usually not presented directly, it may well be that students could achieve a sufficient degree of transfer from course content to enable them to improve their scores significantly. In that event, many instructors would probably agree that one facet of a desirable goal of a beginning psychology course had been met, the goal of developing a more objective way of perceiving and understanding human behavior.

While the KAP Test seems to have promise for evaluating the kinds of goals suggested in this article, as far as college students are concerned, still other investigators may wish to use it, either as it is presently constructed, or as a modified version of their own, with other types of populations, such as adults enrolled in extension classes, or special segments of the non-college population.[4]

FOOTNOTES

1. Clerical assistance for this project was furnished by the Division of General Studies, the Department of Psychology, and the Student Counseling Service, University of Illinois. Thanks are due to numerous colleagues who served as judges of items.
2. The author wishes to express his appreciation to Glen Stice and the Educational Testing Service, Princeton, New Jersey, for granting permission to use the Facts About Science Test in the development of the KAP Test.
3. An initial result of this analysis has revealed that the correlation (r) between precourse KAP Test scores and a verbal scholastic aptitude test was .34 for Scale S and .19 for Scale P.
4. Investigators wishing to use the KAP Test for research purposes may obtain a copy, together with a scoring key, by writing to the author, 236 Student Services Bldg., University of Illinois, Champaign, Illinois. Difficulty level and discrimination values for each item are also available if desired.

REFERENCES

1. Bendig, A.W. and Hountras, P. T. "College Student Stereotypes of the Personality Traits of Research Scientists," Journal of Educational Psychology (1958) ILIX pp. 309-14.
2. Roe, Anne. The Making of a Scientist (1953) New York: Dodd, Mead and Co.
3. Roe, Anne. The Psychology of Occupations (1956) New York: John Wiley.
4. Ross, C.C. and Stanley, J.C. Measurement in Today's Schools. (1954) New York: Prentice Hall.
5. Terman, L.M. "Are Scientists Different?" Scientific American (1955) pp. 192; 25-29.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

THE INFLUENCE OF LIBRARY WORK IN IMPROVING ENGLISH LANGUAGE SKILLS AT THE HIGH SCHOOL LEVEL

DORTHY M. H. HASTINGS
Pacifica High School
Pittsburg, California

DANIEL TANNER
Northwestern University

IS THE FORMAL, systematic, and normative study of English grammar at the high school level the best approach to developing effective English language skills? Surveys of current practices indicate that the prevailing approach among teachers of English at the secondary level is to drill students in the parts of speech and elements of sentence structure without relating this effectively to reading and composition (1). Studies have shown that such approaches to instruction in grammar are of dubious value in improving the functional use of language (2). Evans expresses doubt that formal grammar can be taught to advantage before the senior year in high school (3). Yet, despite the lack of correlation between drill work in formal grammar and developmental ability in English composition and reading many teachers persist in such practices.

Purpose of Study

The purpose of the study was to determine whether improved English language skills could be developed at the tenth-grade level through systematic library experiences in place of the traditional emphasis on formal English grammar. The following hypotheses were tested in the experiment.

1. Instruction and experience in the use of a variety of reference works by an English class will result in higher scores in language skills tests than will be evident for other classes which make little or no use of such materials.
2. Without emphasis on either traditional grammar or spelling, the English class which uses the available library reference works regularly will show equal or better performance on tests involving the mechanics of English and spelling than those which make limited or no use of such works.
3. Experience in library research, involving the use of reference works for the purpose of interpreting information in reports, will enhance the ability of students to ascertain the meaning of words from context.
4. The grammar aspect of language skills, as measured by standardized tests, can be taught outside the traditional classroom setting and in situations vastly different from the traditional emphasis on drill work in grammar.

Methods and Procedures

Subjects of the experiment were four matched high tenth-grade English classes of a large comprehensive high school in San Francisco. The students had been classified by the counseling staff as "middle ability" groups — matched according to data from ability tests, achievement tests, and teacher recommendations.

For purposes of this study, two of the classes were designated as Experimental Groups ("X-1" and "X-2"), while the remaining two classes were designated as Control Groups ("C-1" and "C-2"). The classes ranged in size from 24 to 28 students. All four groups submitted to pretests with the California Language Test, Form W,

to determine whether any significant differences in English language skills were existent prior to initiating the experiment. An analysis of variance for the four groups revealed no statistically significant differences among the groups at the beginning of the experiment.

Each of the four groups was under the tutelage of a different instructor. With the exception of Experimental Group "X-1", each instructor was fully certificated and had a record of several years of successful teaching experience in English language arts at this grade level. The Experimental Group "X-1" was taught by a student teacher. While one might ordinarily assume that a student teacher could not perform as proficiently as an experienced teacher, the results of the experiment, as indicated by the experimental criteria and the testing instruments, failed to verify any such handicap on the part of the student teacher.

All teachers served in the experiment voluntarily and, except for experimental group modifications, proceeded in a pattern closely resembling their usual approach to instruction in English language arts over a thirteen-week period.

Experimental Group "X-1". Experimental Group "X-1", under the direction of the student teacher, devoted one full period per week in the school library where the students engaged in systematic work involving the use of a wide variety of reference materials for information-seeking assignments. The variety of information required of the students in this experimental group was designed to provide some area of interest to all students, however varied their interests might be. No special reports were made as a result of the research. The aim was to acquaint the students with an assortment of reference works to which they might readily refer should the need arise. No emphasis was given to the teaching of formal grammar or spelling except as particular problems arose within the class. Otherwise, the regular class periods during the other four days of the week were devoted to the usual course of instruction for this grade level.

Experimental Group "X-2". Experimental Group "X-2" emphasized using the information gained in the library for the purpose of making reports and completing similar assignments. In addition, considerable time was given to the teaching of traditional grammar and spelling by the column method. Two quizzes and a fifty- to one hundred-word test were administered weekly.

Control Group "C-1". Control Group "C-1" used the library only occasionally and for the purpose of learning to outline various materials selected by the librarian. No library reference works were used by this group. Considerable emphasis was given to instruction in formal grammar and drill work in spelling during the course of instruction.

Control Group "C-2". Control Group "C-2" used the library infrequently and only for the purpose of browsing and selecting books for reading enrichment. Like the other Control Group, considerable emphasis was place on instruction in traditional grammar and drill work in spelling during the regular class periods.

Testing Instruments

As previously stated, the pre-tests revealed no significant differences on the California Language Tests among the four groups, and all four groups had been matched according to the intelligence and prior academic performance. The following pre-tests and post-tests were employed to test the hypotheses of the experiment:

1. Pre-test of all groups using the California Language Test, Form W.
2. Post-test of all groups using the California Language Test, Form X.
3. End-of-course test of all groups using the Durost-Center Word Mastery Test.
4. End-of-course test of all groups using the Iowa Tests of Educational Development, subtest No. 9: Use of Sources of Information.

The California Language Test is divided into two sections: Mechanics of English and Grammar, and Spelling. The Use of Sources of Information section of the Iowa Tests of Educational Development purports to measure familiarity with the contents of standard reference sources in general, some periodical sources in particular, and practical judgment in choosing the proper reference for a particular need. The Durost-Center Word Mastery Test is designed to provide a measure of the general vocabulary level of the secondary school student and to measure the extent to which the student is able to learn the meanings of unknown words by seeing them in typical context situations.

Findings

A comparison between the pre-test and post-test scores in the California Language Test reveals that the aggregate experimental groups made statistically significant gains over their initial scores in total language skills and spelling. On the other hand, the aggregate control groups failed to make significant gains over their initial test scores. See Table I.

TABLE I

GAINS FOR AGGREGATE EXPERIMENTAL GROUPS AND AGGREGATE CONTROL GROUPS FROM PRE-TEST TO POST-TEST ON CALIFORNIA LANGUAGE TEST*

Test	Experimental Group		Control Group	
	t	Level of Significance	t	Level of Significance
Language	1.16	NS	1.64	NS
Spelling	5.59	1%	1.24	NS
Total Language Skills	2.11	5%	1.39	NS

* The t's indicate that there is a significant difference between pre- and post-tests in the experimental group, but not the control group. The language t's are not quite significant.

TABLE II

AGGREGATE EXPERIMENTAL GROUPS VERSUS AGGREGATE CONTROL GROUPS IN PRE-TESTS AND POST-TESTS ON THE CALIFORNIA LANGUAGE TEST*

Test	Pre-Test		Post-test	
	t	Level of Significance	t	Level of Significance
Language	1.11	NS	1.68	NS
Spelling	.86	NS	3.63	1%
Total Language Skills	1.19	NS	2.36	2%

* The t's indicate that there is better than chance difference favoring the experimental over the control groups on the post-test. The language t on the post-test, though significant only at the 10% level, is still higher than on the pre-test for the experimental groups.

Comparisons of the post-test performance on the California Language Test between the aggregate experimental and the aggregate control groups, as shown in Table II, reveals a statistically significant level of attainment favoring the experimental groups over the control groups. It should be remembered that the pre-tests revealed no statistically significant difference between these two groups.

Table III presents a comparison of each experimental and control group according to the results of three standardized tests: California Language Test, Durost-Center Word Mastery Test, and Iowa Test of Educational Development (Use of Sources of Information).

These test-score comparisons reveal that, at the end of the semester, Experimental Group "X-1" outscored every other group at statistically significant levels on the California Language Test and the Iowa Test of Educational Development (Use of Sources of Information). Although Experimental Group "X-2" was regularly the runner-up in these tests, the differences between Experimental Group "X-2" and Control Group "C-1" were not statistically significant. Control Group "C-2" trailed consistently in every test. The results of the Word Mastery Test show that Control Group "C-2" is significantly lower than Control Group "C-1."

It is particularly interesting to note that while Experimental Group "X-1" was taught by a student teacher, each of the other groups was taught by an experienced and fully certificated teacher. While one might assume that the experienced teachers would be more skillful than a student teacher, the results favor the performance of the student teacher's group at statistically significant levels. The question then arises as to whether the enthusiasm of the student teacher created a vector in favor of Experimental Group "X-1." But regardless of the possibility of such disparate variables in instructors among the groups under comparison, the fact remains that the group that avoided all traditional emphasis on formal grammar and spelling actually yielded results superior in these dimensions at statistically significant levels over the remaining three comparison groups.

Summary and Conclusions

The purpose of the research was to determine whether improved English language skills could be developed at the tenth-grade level through systematic library experiences in place of the traditional emphasis on formal English grammar. Four groups of tenth-grade students, matched according to ability and achievement levels, were selected as subjects for the experiment. Pre-tests indicated that all four groups were of comparable level in English language skills. Two of the classes were designated as Experimental Groups ("X-1" and "X-2"), while the remaining two classes were designated as Control Groups ("C-1" and "C-2").

Both experimental groups made systematic use of the school library, whereas the two control groups devoted virtually no time to systematic library work. Experimental Group "X-1" was the only class that eliminated all formal work on grammar and spelling. This group emphasized the use of a wide variety of library references, while Experimental Group "X-2" used the library to gain information for assigned reports.

Despite the fact that Experimental Group "X-1" eliminated the traditional emphasis on formal grammar and spelling, except as particular problems arose within a functional context, this group tended to surpass all other classes at statistically significant levels in tests involving grammar and spelling skills. The statistical data comparing the two experimental and the two control groups reveal that:

a. Experimental Group "X-1" was superior to Experimental Group "X-2" and Control Groups "C-1" and "C-2".
b. Experimental Group "X-2" was superior to Control Group "C-2".
c. Control Group "C-1" was superior to Control Group "C-2", but not statistically significant from Experimental Group "X-2".

As a result of this experiment, the following points of summary and conclusion have been reached:

1. The group that eliminated all traditional emphasis on formal grammar and spelling, and provided systematic work in the use of library references, yielded a level of attainment in English language skills that was statistically superior to the other three comparison groups which emphasized the traditional work in grammar and spelling.

2. The aggregate experimental groups which were given instruction and experience in the use of a variety of library references made significantly higher scores in spelling, mechanics of English, and total language skills tests than did the control groups that made little or no use of such library work.

3. Experience in library work apparently enhanced the ability of students to ascertain the meanings of words from context.

4. The traditional emphasis on formal grammar and spelling does not appear to be the most effective approach to the functional development of such skills for students at the tenth-grade level.

Consequently, the hypotheses of the study were found tenable. Despite the possibility that the quality of instruction may have constituted a disparate variable among the four groups, it has, nevertheless, been shown that the grammar and spelling aspects of English language skills can be taught effectively without the traditionally formal emphasis of drill work in these areas. One might reasonably conclude that is indeed worthwhile for the English teacher to provide systematic experiences in library reference work throughout the course of instruction.

REFERENCES

1. DeBoer, John J. "Grammar in Language Teaching," Elementary English, October 1959, p. 2.
2. Ibid., p. 3.
3. Evans, Bergen. "Grammar for Today," The Atlantic Monthly, CCV (March 1960), pp. 79-82.

TABLE III

COMPARISONS BETWEEN GROUPS ON THREE TESTS OF DEVELOPMENT IN ENGLISH LANGUAGE SKILLS

Test		X-1 vs X-2 t	X-1 vs X-2 p	C-1 vs C-2 t	C-1 vs C-2 p	X-2 vs C-1 t	X-2 vs C-1 p
Language Skills	Pre	-.25	NS	1.63	NS	-1.52	NS
	Post	2.02	5%	2.34	5%	- .93	NS
Spelling	Pre	-.06	NS	-.48	NS	- .33	NS
	Post	3.24	1%	1.72	NS	.22	NS
Total Language	Pre	-.24	NS	1.28	NS	-1.39	NS
	Post	2.47	5%	2.51	5%	- .72	NS
Durost-Center	Pt I	-.54	NS	1.06	NS	- .53	NS
	Pt II	.89	NS	3.03	1%	- .49	NS
	Dif.	1.99	NS	3.29	1%	- .12	NS
Iowa Use		3.28	1%	-.16	NS	.83	NS

p Level of significance
NS Not significant
\- Indicates Mean of X-1 (C-1 or X-2) is smaller

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

LEARNING ABOUT TIME ZONES IN GRADES FOUR, FIVE, AND SIX[1]

O. L. DAVIS, Jr.
Kent State University
Kent, Ohio

INSTRUCTION AND LEARNING in the social studies are highly verbal. Because of this, the development of accurate concepts is difficult, while nevertheless essential. Some authorities in the social studies maintain that too many concepts are presented too soon to pupils. They insist that the immaturity of children does not permit their development of clear concepts and that insufficient time is allotted to the slow process of concept formation.

Children's misconceptions of social studies' concepts long have been noted. Scott and Myers (8) reported many inaccuracies in concepts concerning common terms in history and geography. Confusion in geographic meanings held by elementary school children has been documented further by other studies, many summarized by Davis (1).

Concepts of time and space are two of the many social studies concepts which have received research attention. Location, distance, direction, orientation, and chronology are a few of the meanings involved in these concepts. Ideas concerning geographical time zones are related to concepts of both time and space.

Wesley and Adams, among others, unequivocally hold that learning about geographic time zones is too difficult for children until late in their elementary schooling (10, pp. 307-8). They hold that only the process of maturation will enable children to understand the sphericity of the earth. In programs for the development of the skills of effective map and globe interpretation, understanding geographic time zones is consistently placed in the upper elementary grades, six or above (5; 11).

Little classroom experimentation has been reported which reveals at which age and grade levels concepts of geographic time and space can be learned. Walker (9) found that seven- and eight-year olds could understand simple earth-solar system relationships. While unsubstantiated by statistical evidence, reports by Kelton and Hotchkiss (4) and Schaeffer (7) were enthusiastic about teaching these concepts to fourth graders. Forsyth (2) concluded that children of junior high school age can learn the concept of the sphericity of the earth in addition to certain map reading skills.

The available research, while meager and fragmentary, indicates that growth in the geographic time and space concept does occur and that children may profit from instruction in these concepts. The present study was designed to obtain additional evidence about the effect of instruction on the development of the geographic time and space concept.

Method

The purpose of this experiment was to determine if children in the fourth, fifth, and sixth grades could profit from instruction in concepts of time and space relating to geographic time zones. Five specific null hypotheses were tested. It was predicted there would be no significant interaction and no significant differences between grade levels and between experimental conditions with respect to performances on the criterion measure of time zone learnings:

1. at the conclusion of the experimental teaching period,
2. one month after the conclusion of the experimental teaching period, and with respect to the gain scores,
3. between the administrations of the pre-test and immediate post-test,
4. between the administrations of the pre-test and the test of delayed recall, and
5. between the administrations of the immediate post-test and test of delayed recall.

Subjects

Two classes each of fourth, fifth, and sixth

grades were selected from six different Davidson County, Tennessee, elementary schools. One class at each grade level was designated as the experimental group, the other as the control group.[2] The subjects were white children determined as being members of the middle social class. At the beginning of the experiment, the following numbers of pupils were regularly enrolled in these classes: 4E, 32; 4C, 36; 5E, 34; 5C, 34; 6E, 34; 6C, 36. Because some children in each class were absent on testing days, it was necessary to eliminate their scores from the statistical analysis. Further, a few pupils were randomly eliminated in order to have the same number in each group, twenty-seven. All statistical analyses were made on the data provided by the twenty-seven pupils in each class. Data on chronological age, intelligence, and social studies achievement for the six classes are presented in Table I.

Instruments

Prior to the initiation of the experiment, IQ and social studies achievement scores were obtained. The Lorge-Thorndike Intelligence Test, Level 3, Form B, Nonverbal, was administered to all Ss. The Sequential Tests of Educational Progress Social Studies, Form 4A, was used to assess Ss' achievement in the social studies.

The criterion test employed in this experiment was a Test on Time Zones constructed by the experimenter. This test was composed of forty-six items. Questions related to direction, rotation of the earth, clock time, the International Date Line, and standard time zones in the United States and the world. Administration time was not limited, but all children in all grades finished it within thirty minutes. The reliability of this test was computed by the test-retest method using the scores of the three control classes made on the pre-test and the immediate post-test administrations. This overall correlation coefficient was .83; for grade four, .64; for grade five, .76; for grade six, .84. The test had acceptable content validity as judged by the experimenter and two competent authorities in the social studies.

This criterion measure was administered as a pre-test, an immediate post-test, and as a test of delayed recall. These administrations were designated as TZ1, TZ2, and TZ3, respectively. Data on the classes' performances on the criterion measure are presented in Table II.

Using a 3 x 2 factorial design, analyses of variance were computed for the groups' intelligence quotients, social studies achievement, and performance on the criterion pre-test, TZ1.[3] A summary of these analyses of variance is presented in Table III.

As Table III shows, there was no significant interaction and no significant differences between experimental and control conditions on any of these three measures. There was also no significant difference in IQ between grade levels. Obtained significant differences between grade levels with regard to social studies achievement and performances on the initial criterion test were further analyzed by means of _t_ tests. With respect to social studies achievement, fifth grades scored higher than the fourth grades ($t = 3.07$); the sixth grades scored higher than both the fifth grades ($t = 2.58$), and the fourth grades ($t = 5.65$). On TZ1, likewise, the fifth grades scored higher than the fourth grades ($t = 4.98$) and the sixth grades higher than both the fourth grades ($t = 8.85$) and the fifth grades ($t = 3.87$).

Thus, prior to the experimental teaching period the experimental and control groups did not differ significantly in terms of intelligence, social studies achievement, and understanding of time zones. Differences between school grades on social studies achievement and understanding of time zones were what reasonably should have been expected.

Procedure

The experimenter met with the principals and the teachers of the classes and the intermediate grades supervisor in the second week of January, 1958. The design of the experiment was discussed and a schedule for the testing and experimental teaching was determined. It was agreed that the experimenter would be introduced to all the classes as a student teacher. The teachers of the control classes agreed that they would give their groups no specific instruction regarding geographic time zones during the experimental period except that which was a part of the normal program of study. It was further agreed that all of the pupils would not be told that there were other classes being taught by the experimenter. As far as it could be determined, none of the pupils had had prior instruction in regard to geographic time zones.

The experimental classes were taught a unit specifically embodying material relating to the development of an understanding of geographic time zones. The experimental unit was taught during the classes' regular social studies period. Each experimental group was taught each day for fourteen class days over a period of three weeks. The experimental classes were taught in the sequence of 5E, 4E, 6E. Alternation of times of instruction was not possible because of conflicts in regularly scheduled activities in all three schools. The period of daily instruction was approximately thirty minutes. The experimenter taught the unit to control for variations in teaching method, learning materials, and subject matter presented.

TABLE I

SUMMARY OF DATA ON CHRONOLOGICAL AGE, INTELLIGENCE QUOTIENT AND SOCIAL STUDIES ACHIEVEMENT FOR EXPERIMENTAL AND CONTROL CLASSES

Class	Chronological Age		Intelligence Quotient		Social Studies Achievement	
	Yr.	Mo.	M	SD	M	SD
4E	9	10	101.74	12.18	244.85	6.37
4C	9	9	103.74	12.68	243.81	7.99
5E	10	7	104.04	13.39	253.40	12.00
5C	10	9	105.93	11.53	251.85	11.91
6E	11	8	104.22	14.42	261.89	10.60
6C	11	7	96.78	13.26	257.30	10.95

TABLE II

MEANS AND SDs OF THE THREE ADMINISTRATIONS OF THE CRITERION MEASURE

Class	TZ1		TZ2		TZ3	
	M	SD	M	SD	M	SD
4E	20.93	4.42	33.19	6.24	32.04	8.15
4C	18.52	5.42	20.81	6.03	22.74	5.54
5E	25.74	6.90	37.15	6.68	38.52	5.90
5C	25.96	5.85	27.00	7.23	29.00	6.60
6E	30.22	7.36	39.44	6.70	40.11	5.18
6C	31.00	7.40	31.30	6.70	32.74	7.33

Results

Data obtained from the administrations of the immediate and delayed post-tests and gain scores between the three administrations of the criterion tests were subjected to analyses of variance in a 3 x 2 factorial design. A summary of these analyses appears in Table IV.

Performances on Criterion Tests

As indicated in Table IV, the experimental classes performed significantly better than did the control classes on both TZ2 and TZ3. This indicates that the experimental classes significantly profited from the instruction about geographic time zones. There was no significant interaction. Differences between grade levels observed at the beginning of the experiment persisted. Comparisons of the means are given below. The critical value for the two-tailed test of significance is $t = 2.01$ ($df = 156$).

a. Immediately following the experimental teaching period, the fifth grades scored higher than the fourth grades ($t = 3.99$); the sixth grades scored higher than the fifth grades ($t = 2.60$) and the fourth grades ($t = 6.59$).
b. One month following the conclusion of the experimental teaching period, the fifth grades were still superior to the fourth grades ($t = 7.17$).

Analyses of the Gain Scores

With regard to gains between the administrations of the criterion test, Table IV shows that the experimental classes gained significantly more in understanding of time zones than did the control classes throughout the course of this experiment. There was no significant interaction at any point tested. On only one analysis, that considering the gain between TZ1 and TZ2, was there a significant difference between grade levels. Here, there was no significant difference between the gains of the fourth and the fifth grades ($t = 1.07$) or between the gains of the fifth and sixth grades ($t = 1.67$); the fourth grades indicated a significant gain over the sixth grades ($t = 2.74$). This finding may well have been due to the limitations of the criterion test, the upper limits of which may not have been high enough.

Discussion

The results of this study provide evidence that children, at a younger age than heretofore recognized by certain curriculum authorities, can profit from instruction about geographic time zones. Social studies curriculum theory which advocates deferment of certain concepts, such as those of time and space, has been justified on the basis of the slow maturation of children's concepts and the misconceptions found in children's thinking. If this present study is indicative of possible findings in other areas of concept development, a "deferment-of-instruction" theory needs radical revision.

This study cannot be interpreted as one that minimizes, to any extent, the importance of maturation in the development of concepts. It does indicate that substantial progress in understanding the complexity of geographic time zones can be made by children. A tentative hypothesis is advanced that the limits now believed imposed by maturation on the development of other concepts of time and space may not be as rigid as believed. Results of experimental research will substantiate or refute this hypothesis.

Intermediate grade children in this study learned about direction, sphericity of the earth, rotation of the earth, and earth-sun relationships relating to time, clock time, and time zones. Although outside the scope of this investigation, the investigator believes that some aspects of these concepts profitably can be learned earlier. Effective use can be made of the globe and various map projections, including a polar projection, in teaching these ideas in the intermediate grades. Firm and substantial beginnings can be made in the development of geographic time and space concepts as early as the fourth grade. During each subsequent year's program, provision should be made for additional opportunities to use these learnings in many and varied situations and for additional and progressively more complex learnings about such concepts.

Summary

Two classes each of fourth, fifth, and sixth grades were selected from six different Davidson County, Tennessee, elementary schools. One class at each grade level was designated as the experimental group, the other as the control group. The subjects were middle-class white children. At the beginning of the experiment, there were no significant differences between the experimental and control groups with regard to intelligence, social studies achievement, and understandings about time zones. The experimental classes were taught a unit specifically embodying material relating to the development of an understanding of geographic time zones.

All experimental classes profited from the instruction about geographic time zones. At all points of development observed, sixth graders demonstrated significantly better understanding than the younger children and fifth graders comprehended significantly more than the fourth

TABLE III

SUMMARY OF ANALYSES OF VARIANCE OF IQ, SOCIAL STUDIES ACHIEVEMENT, AND SCORES ON TZ1

Variable	Variance Components for the Several Analyses			
	Between Conditions df=1	Between Grade Levels df=2	Interaction df=2	Within df=156
Intelligence Quotient	56.89	271.13	396.72	167.52
Social Studies Achievement	232.33	3151.41* (F=30.16)	49.80	104.49
Scores on TZ1	8.92	1609.12* (F=39.09)	39.08	41.16

* The ratio between each of these variance components and its appropriate error term is significant at or beyond the five percent level.

TABLE IV

SUMMARY OF ANALYSES OF VARIANCE OF PERFORMANCES ON AND GAIN SCORES BETWEEN ADMINISTRATIONS OF CRITERION TEST

Variable	Variance Components for the Several Analyses			
	Between Conditions df=1	Between Grade Levels df=2	Interaction df=2	Within df=156
Performance on TZ2	4234.00* (F=96.84)	960.07* (F=21.97)	60.22	43.70
Performance on TZ3	3085.49* (F=72.26)	1164.24* (F=27.27)	18.84	42.70
Gain Between TZ1 and TZ2+	3755.55* (F=141.93)	100.26* (F= 3.79)	12.24	26.46
Gain Between TZ1 and TZ3+	2648.30* (F=90.76)	88.02	25.64	29.18
Gain Between TZ2 and TZ3+	93.39* (F= 5.08)	24.02	25.40	18.39

* The ratio between each of these variance components and its appropriate error term is significant at or beyond the five percent level.

\+ A constant of 20 points was added to each true gain score for ease in calculation.

graders. The finding that fourth graders gained significantly more understanding as a result of instruction than did the sixth graders may have been spurious. All experimental classes gained significantly and equally well from the initiation of the experimental teaching period to the test of delayed recall. Too, they continued their significant gain of understanding following termination of instruction.

Interpretation of the results of this experiment raise serious questions about theories of instruction concerning the particular concepts considered in this study. Children may be able to profit from instruction in other geographic concepts earlier than thought possible. Questions are raised about the appropriateness of deferred instruction about other concepts. Much additional experimental evidence is needed about concept development in the social studies.

FOOTNOTES

1. This study was done at the George Peabody College for Teachers as part of a doctoral dissertation under the supervision of Harold D. Drummond.
2. All future references to these classes will be made in terms of their grade and experimental condition, i.e., Fourth Control, Fourth Experimental, Fifth Control, etc. At times, these references may be abbreviated, i.e., 4C, 4E, 5C, etc.
3. All data obtained in this experiment were treated by analysis of variance and, where appropriate, *t* tests. Two-tailed tests were employed and the five percent level of significance was used throughout. Hartley's test (3) was used to test for homogeneity of variance before proceeding with each analysis. In instances when the hypothesis of homogeneity of variance was not accepted, the departure was not so extreme as to have had appreciable effect on the assumption underlying the F-test (6, p. 86).

REFERENCES

1. Davis, O.L., Jr. *Learning about Time Zones: An Experiment in the Development of Certain Time and Space Concepts*, unpublished Ph.D. dissertation, George Peabody College for Teachers, 1958.
2. Forsyth, Elaine. *An Experiment in the Teaching of Certain Map Reading Skills at the Junior High School Level*, unpublished Ph.D. dissertation, Cornell University, 1943.
3. Hartley, H.O. "The Maximum F-ratio as a Short-cut Test of Heterogeneity of Variance," *Biometrika*, XXXVII (1950), pp. 308-12.
4. Kelton, Mary E., and Hotchkiss, Caroline W. "An Experiment in Fourth Grade Geography," *Journal of Geography*, XIV (1916), pp. 245-51.
5. Kohn, C.F., et al. "Interpreting Maps and Globes," in Helen M. Carpenter, ed., *Skills in Social Studies*. *Twenty-fourth Yearbook of the National Council for the Social Studies* (Washington, D.C.: The Council, 1953).
6. Lindquist, E.F. *Design and Analysis of Experiments in Psychology and Education* (Boston: Houghton Mifflin, 1953).
7. Schaeffler, Grace C. "An Informational Unit on Time," *Elementary School Journal*, XXXVIII (1937), pp. 114-17.
8. Scott, Flora, and Myers, G.C. "Children's Empty and Erroneous Concepts of the Commonplace," *Journal of Educational Research*, VIII (1923), pp. 327-34.
9. Walker, Helen E. *Selected Time-Space Concepts of Seven- and Eight-Year-Old Children*, unpublished M.S. in Education thesis. New Jersey State Teachers College, Glassboro, 1952.
10. Wesley, E.B., and Adams, Mary A. *Teaching Social Studies in Elementary Schools*, revised edition (Boston: Heath, 1952).
11. Whipple, Gertrude. *How to Introduce Maps and Globes*, No. 15: *How To Do It Series* (Washington: National Council for the Social Studies, 1953).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

A PILOT STUDY TO DETERMINE THE RELATIVE IMPORTANCE OF SELECTED PERSONAL AND PROFESSIONAL FACTORS IN THE SUCCESS OF THE STUDENT TEACHER IN SCIENCE

JOHN GABRIEL NAVARRA and RUTH ROBERTA DUGAN
Jersey City State College

Purpose for the Study

This investigation was designed to make a study of selected personal and professional factors in the background of nineteen senior science majors at Jersey City State College in an effort to explore possible factors predictive of success in science teaching.

The selection of the personal and professional factors studied was based on the following criteria:

1. The item has been traditionally accepted by the public, teachers, or administrators as important in teacher selection.
   Example: College Scholastic Average, IQ, National Teacher Examination Scores.
2. The items have been found to be significant for academic or professional success by other studies.
   Example: Professional Viewpoints (1), Reasons for Vocational Choice (2), good mental health (3;4), teacher attitudes (5).
3. The item has been considered important to success because of the value judgments of experts.
   Example: High service goals, unselfishness, patience, professional training, leadership qualities, democratic behavior, ability to express thoughts clearly and effectively (6;7;8;9).

Since other investigations have revealed the difficulties involved in defining and determining success (10;11;12), eight kinds of criteria were used in this study in the attempt to discriminate between the effective and ineffective student science teachers:

1. Mean Scores as Evaluated by College Supervisors on a Critical Teachers Behavior Scale.

This scale is an incorporation of the two scales found useful in the Teacher Characteristics Studies under the direction of Dr. Ryans (13) and sponsored by the American Council on Education. This scale describes critical behaviors of teachers that produce or retard (according to the judgment of the investigators) behavior changes in students in the direction of the following educational objectives: the development of skills, understandings, work habits, desirable attitudes, and personal adjustment.

2. Mean Scores on the Critical Teachers Behavior Scale as evaluated by the Cooperating Teacher under whom the student did his student teaching.

3. Student self-rating on a Critical Teachers Behavior Scale.

4. Mean Scores as evaluated by Science Faculty on Science Student Behavior Scale.

Science Student Behavior Scale was designed by members of a committee on evaluation in the

Science Department at Jersey City State College and describes behavior changes in students in the direction of the following educational objectives: development of critical thinking, development of scientific attitudes, personal adjustment, transfer of training, development of skills, understandings, and work habits.

5. National Teacher Examination Scores.

6. College Scholastic Average.

7. Mean scores of student rank according to potential effectiveness as a science teacher as judged by science faculty.

The names of the science student teachers in the study were listed and the following directions were given to the faculty members evaluating them: Kindly rank the following prospective science teachers according to your estimate of their potential ability to teach effectively. Give a rank of one to your first choice, two to your second, etc, and 19 to the person you feel would be least effective among the people listed.

8. Mean scores of student rank according to potential effectiveness as a science teacher as judged by peers.

The directions to peers were exactly as described in the above paragraph with the addition of one sentence: If you are a Science major participating in the investigation, do <u>not</u> rank yourself.

<u>Procedure in Collecting Data</u>

Nineteen senior science majors who had completed their student teaching and who had been together as a group since freshman year participated in the collection of the following data:

1. Personal Factors
   a. IQ scores as determined by the Otis Self-Administering Test of Mental Ability, Form A.
   b. Mental Objectivity - Egocentricity, Extroversion - Introversion proportion scores as determined by two tests: Dugan Teacher Action Inventory (14) and Miller Murray Self Perception Inventory (15).

Mental objectivity as defined by Miller and Murray is the ability of a person to take into consideration what is going on in the thinking of others about him. The mentally objective person, state Miller and Murray:

. . . governs his speech and behavior accordingly to achieve and maintain smooth and warm social contacts. He is able to concentrate completely on the work in hand without undue regard to self . . . is well oriented toward physical and social reality . . . sees more degrees of relationships in a situation, avoids dogmatism, hasty generalizations, 'allness' attitudes and is more free of prejudices . . . . Objectivity provides a basis for intergrativeness and creativity in human relationships and enterprises (16).

Egocentricity as defined by Miller and Murray, is the failure to take into consideration what is going on in the thinking or evaluation of other persons in the situation. It is a preoccupation with self or with the actions of others in relation to the self which prevents appropriate or adequate human relations (17). In their description of the egocentric person, Miller and Murray observe:

. . . he is characterized by either or both (1) emotional interferences or resistances to concentration and control, (2) disregard, indifference or intolerance of the thinking and feelings of others. Egocentricity, typically, is best represented in the human relations and speech of children and emotionally immature adults . . . because egocentric behavior represents types of reaction reflecting less dominance of the 'cortex' or intellect and too much 'thalamus': this person does not discriminate closely, has poor perspectives, misvaluates and consequently has poor predictability (18).

Extroversion is the tendency on the part of an individual to react often in various situations in the following manner: The extroverted person projects what is going on within himself upon the outside world, dislikes doing work requiring detailed attention, makes decisions quickly, makes friends readily, acts on hunches rather than logic, is carefree about possessions, works well with other people, and is enthusiastic.

Introversion is the tendency on the part of an individual to react often in various situations in the following manner: The introverted person keeps what is going on within himself to himself, likes to work with details, acts slowly and logically, makes friends slowly, is orderly and meticulous about possessions, dislikes rules and regulations, and prefers to work alone.

   c. Life Goals as Listed by Cooperative Tests Life Goals Scale.

2. Professional Factors
   a. Reasons for Vocational Choice Found to be a factor worthy of further study in an investigation of teacher success by Ringness (19).
   b. Professional Viewpoints as determined by a scale designed using Meyer's (20)

analysis of the conflicts between traditional and progressive educational philosophy.

c. Mean Scores of student rank according to potential effectiveness as a science teacher as judged by peers.
d. Student self-rating on Critical Teachers Behavior Scale.

Administration and Science Faculty participated in the collection of the following professional data:

e. Mean Scores as evaluated by College Supervisors on the Critical Teachers Behavior Scale.
f. Mean Scores as evaluated by Cooperating Teacher on The Critical Teachers Behavior Scale.
g. Mean Scores as evaluated by Science Faculty on Science Student Behavior Scale.
h. Mean Scores of student rank according to potential effectiveness as a science teacher as judged by Science Faculty.
i. College scholastic average.
j. National Teacher Examination scores.

Procedure and Results in Treating Data

1. To obtain one numerical evaluation which would reflect a multidimensional appraisal of each student teacher's effectiveness, the following scores were converted to rank scores and the mean of these rank scores were derived. This score is called the Multidimensional Appraisal score. See Table I.
   a. National Teacher Examination Scores.
   b. Mean Scores as evaluated by Science Faculty on Science Student Behavior Scale.
   c. College Scholastic Average.
   d. Mean Scores of student rank according to potential effectiveness as a science teacher as judged by Science Faculty.
   e. Mean Scores of student rank according to potential effectiveness as a science teacher as judged by peers.

Results: The range of the Multidimensional Appraisal Scores was from 2.8 (most effective) to 15.2 (least effective). $M = 8.8 \pm .82$; $\sigma = 3.46 \pm .56$.

2. Examination of raw data of Miller Murray Self Perception Inventory Scores (21) and Dugan Teacher Action Inventory Scores (22) confirmed previous findings of high correlation ($r = .80$; $t = 10.88$; $p < .01$) between the two inventories. Therefore, only the egocentricity and extroversion scores on the Dugan Teacher Action Inventory were selected for study in this investigation. (According to the structure of the Teacher Action Inventory, mental objectivity and egocentricity scores are always proportions of the whole and added together equal one (1.000); therefore, only one of these factors need be treated statistically. The same is true for the factors, extroversion - introversion.)

Results: Egocentricity $M = .263 \pm .016$; $\sigma = .069 \pm .011$.
Extroversion $M = .514 \pm .008$; $\sigma = .033 \pm .005$.

3. Although the Professional Viewpoint Scale had a possible range of from one (most progressive viewpoint) to 7 (most traditional viewpoint), the actual range of mean scores was between 2.2 and 4.3; $M = 3.1 \pm .15$; $\sigma = .635 \pm .11$.

4. IQ scores on the Otis Self-Administering Test of Mental Ability, Higher Form A, ranged from 95 to 129. $M = 118.5 \pm 2.2$; $\sigma = 9.1 \pm 1.47$.

5. Mean scores of student rank according to potential effectiveness as a science teacher as judged by peers ranged between 2.50 to 14.10. $M = 9.64 \pm .71$; $\sigma = 2.99 \pm .49$.

6. Mean scores of student rank according to potential effectiveness as a science teacher as judged by Science Faculty ranged between 1.8 to 16.0. $M = 9.08 \pm .93$; $\sigma = 3.88 \pm .63$.

7. Mean scores derived from ratings by supervisors and cooperating teachers on The Critical Teachers Behavior Scale were of little value in discriminating between the effective and ineffective practice teachers. Although the scale had a possible range of from one to 7, the scores were skewed in the positive direction toward the effectiveness pole.

Mean scores as rated by Cooperating Teachers ranged from 1.1 to 5.6. $M = 2.6 \pm .58$; $\sigma = 1.91 \pm .39$.

Mean scores as rated by College Supervisors ranged from 1.2 to 4.4. $M = 2.9 \pm .32$; $\sigma = 1.06 \pm .22$.

Since Cooperating Teachers scores and Supervisors scores seemed to have little agreement, they were tested for correlation. Since an r of $+.203 \pm .087$ was obtained, the relationship between these ratings is negligible.

TABLE I

PERSONAL AND PROFESSIONAL FACTORS SCORES

Student Code No.	Mean Score Professional Viewpoint Scale	Egocentricity Teacher Action Inventory	IQ Score	Mean Rank Score Faculty	Mean Rank Score Peers	Multidimensional Appraisal Score
1	3.2	.219	95	7.6	7.75	9.2
2	4.2	.295	122	15.2	13.83	12.0
3	2.6	.196	125	3.5	8.83	5.4
4	absent	.311	121	16.0	13.91	15.2
5	3.4	.343	129	7.0	14.10	8.0
6	2.5	.166	115	12.5	9.33	11.4
7	2.7	.279	113	2.8	5.92	4.8
8	2.2	.098	116	2.4	6.54	5.6
9	2.9	.202	129	1.8	11.38	4.4
10	3.6	.273	125	11.6	12.00	9.6
11	3.9	.345	129	6.8	4.69	4.5
12	2.2	.240	104	8.6	12.33	13.2
13	2.8	.425	115	5.8	10.63	10.6
14	3.2	.340	123	9.4	8.82	9.8
15	4.3	.328	104	12.5	11.38	13.8
16	3.4	.238	116	7.5	8.17	8.2
17	2.6	.161	123	2.3	2.50	2.8
18	3.8	.329	125	15.0	7.17	9.4
19	2.6	.200	125	7.0	9.75	9.8
	M = 3.1 ± .15 σ = .635 ± .11	M = .263 ± .016 σ = .069 ± .011	M = 118.6 ± 2.2 σ = 9.1 ± 1.47	M = 9.08 ± .93 σ = 3.88 ± .63	M = 9.64 ± .71 σ = 2.99 ± .49	M = 8.8 ± .82 σ = 3.46 ± .561

8. Life Goals mentioned as first or second in importance were tallied for all students and percentages of responses were derived as follows:

Desire for Self-Development	21 %
Desire to Serve God	15.8 %
Desire to "Get Ahead"	13.2 %
Desire to Serve Community	10.5 %
Desire to Obtain Lasting Pleasures	7.9 %
Desire to be Able to Handle Problems as They Come	5.3 %
Desire for Peace of Mind	5.3 %
Desire for Security	5.3 %
Desire for Fine Relations with Others	5.3 %
Desire to go to Heaven	5.3 %
Miscellaneous	5.3 %

9. Reasons for Vocational Choice mentioned as first or second in importance were tallied for all students and the percentages of responses were derived as follows:

Variety	18.9 %
Opportunity	18.8 %
Prestige	16.2 %
Security	8.1 %
Serve Society	8.1 %
Ease of Obtaining Education	8.1 %
Second Choice	8.1 %
Clean Surroundings	5.4 %
Family Pressures	2.7 %
Love of Children	2.7 %
Co-Workers	2.7 %

10. Correlation Techniques.

To determine the relative importance of the selected personal and professional factors to the various criteria used in an attempt to discriminate between effectiveness and ineffectiveness in teaching, comparisons were made between various sets of scores by tests of significant correlation.

According to Garrett (23), an r of .40 and above denotes substantial or marked relationship between two tests, and so it has been assumed in this investigation that no significant correlation exists between the following factors unless r = .4 or higher.

A. The Multidimensional Appraisal Score was compared with the following:
   1. Egocentricity scores on Teacher Action Inventory.
      Results: r = .14 + .05
   2. Professional Viewpoints Score.
      Results: r = .26 + .06
   3. Extroversion Scores on Teacher Action Inventory.
      Results: r = .049 + .055
   4. IQ Scores as derived by the Otis Self-Administering Test of Mental Ability, Higher Form A.
      Results: r = -.377 + .048

   A relationship probably exists between these two factors although the r does not meet the test for significance. When IQ is high, the Multidimensional Appraisal Score is low (Effectiveness).

B. Raw data examination suggested further study of the relationship between Egocentricity Scores on the Teacher Action Inventory and the Mean Scores of student rank according to potential effectiveness as a science teacher as judged by peers and as judged by Science Faculty.
   1. Egocentricity Scores and Mean Scores of student rank as judged by peers.
      Results: r = .42 + .046
   2. Egocentricity Scores and Mean Scores of student rank as judged by faculty.
      Results: r = .413 + .046

   A significant relationship exists between a lack of egocentricity (mental - objectivity) and low rank (effectiveness) when faculty members and peers are asked to make a forced choice among a group of students they know well as to potential effectiveness as a science teacher.

C. Raw data suggested agreement between Science Faculty and peer judgments as to student rank according to potential effectiveness as a science teacher.
   Results: r = .434 + .048

   A significant correlation exists between Mean Scores of student rank according to potential effectiveness as a science teacher as judged by peers and as judged by Science Faculty.

D. Since high scores on the Professional Viewpoint Scale represents the traditional or authoritarian and strict discipline school of educational philosophy, it was of interest to investigate the relationship between these scores and scores for egocentricity as measured by the Teacher Action Inventory. According to definition of egocentricity "allness" attitudes and dogmatism are characteristics of the egocentric individual.
   Results: r = .699 + .03

   A significant correlation exists between high scores on the Professional Viewpoint Inventory (traditional philosophy) and high egocentricity scores.

E. No significant correlation existed between Professional Viewpoint Scale Scores and Mean Scores of student rank according to

potential effectiveness as a science teacher as judged by peers.
Results: r = .23 ± .06

F. Significant correlation was found, however, to exist between Professional Viewpoint Scale Scores and Mean Scores of student rank according to potential effectiveness as a science teacher as judged by faculty members.
Results: r = .64 ± .035

Conclusions

1. Discriminations between effective and ineffective student teachers showed reliability only when judges (faculty and peers) were forced to make a choice bv rank.

2. High scores for the factor egocentricity showed significant positive correlation with judged (by both faculty and peers) potential ineffectiveness as a science teacher.

3. High scores for the factor egocentricity showed significant positive correlation with "traditional" educational viewpoints.

4. Significant correlation was found to exist between professional viewpoints and judged potential effectiveness (by faculty members) as a science teacher.

5. Service goals as revealed by Life Goals and Reasons for Vocational Choice questionnaires were not of primary importance to the majority of student teachers in the investigation.

6. A relationship (just below the level of significance) existed between IQ scores and Multidimensional Appraisal of effectiveness when judgments of supervisors, faculty, and students were combined with college scholastic averages and National Teacher Examination Scores.

Some Implications

This initial pilot study has provided a number of leads and checkpoints for the science faculty with respect to the qualitative judgments they make about students. It has also provided a reference point to contrast faculty judgments with student judgments.

A number of anecdotal case studies are being built as a result of and in relation to the data. In addition, some attention is being directed to the situations in which faculty judgments and peer judgments are diametrically opposed.

1. A sensitivity has grown on the part of the staff with respect to students who receive a consistently high rating from the staff, but receive a consistently low rating by the peer group. A student of this type is very interesting. He is never really a part of his peer group. And yet he exerts a tremendous amount of influence on the group. In addition, such a student can move into a high school faculty and become a very active leader. The suspicion is that the success is really based on the same kind of appeal and attractiveness which the student had for the college faculty. For the most part, the student has for his colleagues on the high school faculty an older group than the peer group in college. This kind of situation raises many interesting problems related to success and effectiveness.

2. At times, students are rated consistently low by faculty and yet receive a consistently high rating by the peer group. In our experience, students in this group have been of two types:

a. The first type is the student of low ability as judged by the faculty and ordinary standards of academic achievement. But this student has motivation and a great deal of drive and will to succeed. He is, in fact, a plodder. The peer group seems to be particularly sensitive and responsive to the hard worker. Under most ordinary circumstances this will to succeed by hard work "pays off" in the high school and the student is successful. In such circumstances, the peer judgment has been a better indicator in predicting success.

b. The second type of student is the one with high potential in terms of IQ and other such measures of ability. However, the fondness for group approval and the drive to be part of the group overshadows his willingness to outshine the group. The college faculty seems to be particularly intolerant of such individuals. The high peer judgment with respect to such individuals does not seem to be based on merely popularity since other more popular individuals do not receive a similarly high rating. There is, in fact, an implicit recognition by the peer group of the dormant ability. In most instances within our experience, this ability becomes active when placed in the role of the teacher. The relationship changes and the individual seems to feel that success in this new role does not threaten his relationship with the peer group. It would almost seem that this individual feels that outstanding success in a college course and receipt of a high grade deprives others of a chance to succeed in that same course. This individual seems to have a highly developed sense of loyalty to the peer group.

The three cases cited above are simply initial probes in an attempt to look at the data and supply reasons behind the situations noted. This pilot study has been used to probe, pinpoint, and clarify staff judgments. The problem is really one of gaining experience and insight with respect to some of the personal and professional factors which might be indicative of future success. An extensive follow-up of these graduates is being planned. In addition, the same kind of data is being collected on current graduates.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

RELATIONSHIPS BETWEEN WRITING REQUIRED IN HIGH SCHOOL AND ENGLISH PROFICIENCY IN COLLEGE

ROBERT McQUEEN
A. KEITH MURRAY
FREDERIKA EVANS
University of Nevada

ONE OF THE PRINCIPAL problems facing colleges and universities in admitting freshman students concerns the frequency with which these students display inadequate skills in the use of English. Institutions of higher learning typically deal with such students by placing them in non-credit classes featuring remedial instruction in English. In a recent study carried out by the National Council of Teachers of English (1961), it was estimated that in most institutions 20 percent of college freshmen failed the English placement examination and, in some cases, the rate of failure rose as high as 50 percent.

In undertaking the task of developing basic English skills for their freshman students, colleges and universities do so with much grumbling and protestation. The complaint is made that such instruction is not the proper function of colleges, but, rather, the responsibility of the high schools.

While some high schools have been unwilling to introduce innovation into the English curriculum, McAllister (1953) found much in the way of variation in high school English programs. Some, however, (Morrison, 1951; Franchere, 1952; Pooley and Searles, 1956) have found very little in the way of objective data indicating the apparent superiority of one instructional program over others. It was the purpose of the present investigation to assemble a portion of that objective data.

Problem

The hypothesis advanced in this investigation was that, given groups of students with equivalent academic ability, a positive relationship would obtain between the amount of writing required of groups in high school and the English proficiencies they demonstrate in college. Essentially, the problem was to measure objective criteria of skills associated with the college English proficiency of different high school groups and compare these with the amount of writing that was required of the groups throughout their high school years.

Procedure

The University of Nevada, where the present study was completed, has an undergraduate enrollment of which 80 percent are Nevada high school graduates. Of that large proportion the majority are graduates of one of five high schools. Subjects for this study consisted of all the graduates of these high schools, grouped by high school, who entered the University of Nevada during the four-year period 1957-1960.

Four separate measures were taken on each subject, with the data being pooled with that of other students who entered the university from the same high school. The measures were as follows:

1. the standard composite scores earned on the A. C. E. or A. C. T. college entrance examinations,
2. standard scores earned on the English Department placement test,
3. the frequency with which students from the different high schools were assigned as incoming freshmen to sections of (a) remedial English, (b) regular English, or (c) Honors English, and
4. grades earned at the completion of freshman English classes carrying university credit.

A fifth measure was taken which did not involve assessment of individual subjects, but rather an examination of an aspect of the English curriculum in the five high schools from which the subjects had graduated. Using a questionnaire adapted from one employed by Hook (1961) in a study of award-winning high schools, each English teacher in the high schools involved in this study was interviewed. To reduce ambiguities and otherwise increase the validity of the instrument, the questionnaire was executed by one of the investigators during carefully structured face-to-face interviews. While this procedure necessitated

TABLE I

SUMMARIES OF ANALYSES OF VARIANCE

(1) Standard Composite Scores of A. C. E. and A. C. T. Examinations

Source of Variation	SS	df	MS	F
Between groups	122	4	30. 5	1. 60 P=ns
Within groups	14925	786	18. 99	
Total	15047	790		

(2) Standard Scores of English Placement Test

Source of Variation	SS	df	MS	F
Between groups	639	4	159. 75	8. 10 P=. 01
Within groups	13634	691	19. 73	
Total	14273	695		

(3) Grades Earned in Freshman English Classes

Source of Variation	SS	df	MS	F
Between groups	10. 36	4	2. 59	2. 44 P=. 05
Within groups	717. 06	679	1. 06	
Total	727. 42	683		

TABLE II

ANALYSES OF SIGNIFICANT Fs USING DUNCAN'S MULTIPLE RANGE TEST

(1) Standard Scores of English Placement Test

High School	A	B	C	D	E
Mean	17. 82	18. 75	19. 23	20. 48	21. 67
Number	38	127	64	422	45

Range	Minimum Significant Range
E-A = 24. 79 P= . 05	13. 76
E-B = 23. 80 P= . 05	13. 32
E-C = 17. 74 P= . 05	12. 88
E-D = 10. 73 P= ns	12. 43
D-A = 22. 21 P= . 05	12. 43
D-B = 24. 15 P= . 05	13. 32
D-C = 13. 16 P= . 05	12. 88
C-A = 9. 77 P= ns	12. 43
C-B = 4. 42 P= ns	12. 88
B-A = 7. 07 P= ns	12. 43
	12. 43

(2) Grades Earned in Freshman English Classes

High School	A	B	C	D	E
Mean	1. 80	1. 87	1. 97	2. 13	2. 24
Number	40	102	64	429	49

Range	Minimum Significant Range
E-A = 2. 92 P= ns	3. 19
E-B = 3. 01 P= ns	3. 09
E-C = 2. 02 P= ns	2. 99
E-D = 1. 03 P= ns	2. 88
D-A = 2. 82 P= ns	3. 09
D-B = 3. 34 P= . 05	2. 99
D-C = 1. 68 P= ns	2. 88
C-A = 1. 19 P= ns	2. 99
C-B = 0. 88 P= ns	2. 88
B-A = 0. 53 P= ns	2. 88
	2. 88

extensive travel throughout the state and many hours of interviewing, it was felt that the data thereby gained more than justified its use. During the course of these interviews each teacher was asked to reflect upon his various English classes, referring if necessary to syllabi and assignment records, to ascertain the amount of writing each required in a typical week. Such writing, expressed in number of words written, was limited to those themes and essays submitted for and returned with teacher criticism. Moreover, because individual teachers many times taught more than one class, often at different grade levels, a separate questionnaire was executed for each English class in each high school. The resulting data was, of course, crucial to the present investigation, since it formed the independent variable with which all other measures were compared.

Results and Discussion

The results of this study are presented in Tables I-IV. Table I shows the summaries of analysis of variance procedures dealing with data emerging from A. C. E. and A. C. T. examinations, English Placement Test scores, and grades earned in Freshman English classes. Bartlett's test of homogeneity of variance was run on all three sets of data. With four degrees of freedom a Chi Square, in order to be significant, would have to be 9.49 or greater. All Bartlett's tests fell considerably short of that value. The resulting Chi Square in the case of A. C. E. and A. C. T. scores was 1.50, that for the English Placement Test scores was 4.89, and for grades earned in English classes the Chi Square value was 3.36. The failure of any of the Bartlett's tests to reach statistical significance strengthens the use of analysis of variance in this study.

Part (1) of Table I reflects the analysis of A. C. E. and A. C. T. scores earned by the five high school groups. It will be noted that the resulting F ratio was 1.80 and not statistically significant. The meaning of this statistic, however, has a special importance for this investigation. It signifies that the five groups under study were not significantly different in general college aptitude, at least insofar as the A. C. E. and A. C. T. are able to permit such an estimate. While not shown in the tables, the range of means for these scores was quite small, running from 18.50 in the case of school B to 19.75 in the case of school A. Had group differences in measured college aptitude proved statistically significant, then subsequent group differences in college performance could have been attributed to the differences initially present when the groups entered the university. Since evidently the groups were not different in college aptitude when they enrolled in college, the explanation for observed differences would seem to lie elsewhere.

Part (2) of Table I deals with scores earned on the English Placement test. The F shown is 8.10, which is significant at the .01 level of confidence. This statistic signifies that the observed differences between the five groups on this common criterion were so large that they could not be attributed to chance. For a closer analysis of these differences, attention must be directed to Table II. Before turning to that consideration, however, a final reference to Table I is necessary. It will be noted that the F ratio growing out of the analysis of variance for grades earned in freshman English classes is 2.44, significant at the .05 level of confidence. This statistic means that the observed differences between the five high school groups with respect to class grades was of a sufficient magnitude as to rule out an explanation based upon chance.

Table II presents group-by-group comparisons of the data relating to English Placement Test scores and English class grades, both of which resulted in significant F ratios as shown in Table I. The most appropriate statistic for these comparisons is Duncan's multiple range test described in some detail by Kramer (1956). With respect to group differences in scores on the English Placement Test, shown in part (1) of Table II, it can be seen that the group from school E significantly out-performed school groups A, B, and C. School D, which ranked second to school E, also performed significantly better than the groups from schools A, B, and C. Part (2) of Table II presents group differences in the grades earned in freshman English classes. While only the difference between schools D and B proved statistically significant, it is interesting to note that the ranking in mean grade point averages for the five schools is identical to the ranking which resulted from calculating the mean standard scores earned on the English Placement Test.

Table III shows the frequencies with which students from each high school were assigned to remedial, regular, or honors English classes, together with a Chi Square analysis of these occurrences. Since the overall Chi Square proved significant at the .01 level of confidence, school-by-school Chi Square procedures were calculated. School E, which had the smallest proportion of remedial students and the largest proportion of honors students, was significantly different from all other schools. It will be recalled, at this point, that students from school E were also the group with the highest mean performance on the English Placement Test, as well as the one showing the highest mean grade point average in English classes. School D, once again the second ranking school, was significantly different from school B, due primarily to having relatively fewer students placed in remedial classes and more placed in classes of honors English. The remaining schools were not significantly different, one from the

TABLE III

NUMBER OF STUDENTS FROM EACH HIGH SCHOOL ASSIGNED TO REMEDIAL, REGULAR, AND HONORS ENGLISH CLASSES WITH CHI SQUARE ANALYSES

English Class	High School A	B	C	D	E
Remedial	8	31	16	57	5
Regular	40	105	61	391	37
Honors	2	5	4	48	13
Totals	50	141	81	496	55

Overall Chi Square = 34.01 P = .01

Chi Squares by Schools

E-A = 8.64 P= .02	D-B = 14.00 P= .001
E-B = 21.17 P= .001	D-C = 5.59 P= ns
E-C = 11.88 P= .01	C-A = 0.35 P= ns
E-D = 9.87 P= .01	C-B = 0.35 P= ns
D-A = 2.40 P= ns	B-A = 0.82 P= ns

TABLE IV

NUMBER OF CLASSES IN EACH HIGH SCHOOL REQUIRING EITHER 250 OR MORE THAN 250 WRITTEN WORDS PER WEEK FOR TEACHER CRITICISM WITH CHI SQUARE ANALYSES

Words per week	High School A	B	C	D	E
250 or fewer	12	21	19	33	6
More than 250	10	6	3	24	10
Totals	22	27	22	57	16

Overall Chi Square = 13.48 P= .01

Chi Squares by Schools

E-A = 1.10 P= ns	D-B = 3.03 P= ns
E-B = 6.98 P= .01	D-C = 5.72 P= .05
E-C = 9.85 P= .01	C-A = 5.36 P= .05
E-D = 4.29 P= .05	C-B = 0.60 P= ns
D-A = 0.07 P= ns	B-A = 2.29 P= ns

others, in their assignments to freshman English classes.

Table IV presents a simple summary of the number of classes in each high school which required students to write fewer than 250 words per week and the number of classes in which more than 250 written words were required each week. An overall Chi Square on all these frequencies was significant at the .01 level of confidence, and on the strength of this statistic, school-by-school Chi Squares were then calculated. It will be noted that school E differed significantly from schools B, C, and D. Both schools D and A required significantly more writing than was required of students from school C.

Clearly, school E set far greater writing requirements for its students than was the case in any other school. Again it will be recalled that the E group was the one which consistently out-performed all other schools in the three separate measures of English proficiency. Recalling that no significant differences were found between groups with respect to college aptitude, the sharp differences noted in their high school writing experiences takes on added significance. Indeed, the relationship between the amount of high school writing and college English proficiency appears so strong in this study that to deny the former a principal role in the latter would be to ignore the obvious. While it is unquestionably possible to dissect the high school English curriculum into many parts, emphases, and phases, it is also true that theme and essay writing, submitted for teacher criticism, sets the stage for a great variety of instructional opportunities in English. Moreover, evidence presented in this study strongly suggests that extensive writing in high school best prepares the student for the English program he will later face in college.

Summary

The present study undertook to establish relationships between the amount of theme and essay writing required in five high schools and the college English proficiency of their respective graduates. A test of college aptitude revealed no statistically significant differences between the several groups. In measures of English proficiency, however, including performance on an English Placement Test, assignment to remedial, regular, and honors sections of freshman English, and grades earned in English classes, two high school groups consistently out-performed the others. In fact, the rankings by school in each measure of English proficiency remained identical throughout the study with but one minor exception.

When the amount of writing required in high school was included in the analyses of actual English skills, the high performing group clearly emerged as the one with the greater experience in writing. For the remaining schools, too, there tended to be a positive relationship between demonstrated skills in English and the quantity of theme writing required during high school years. The investigators conclude that a clear relationship between the two has been statistically demonstrated.

REFERENCES

1. Franchere, H. C. "College Freshmen Reconsider and Suggest," College English, (1952) XIII (6), pp. 326-30.

2. McAllister, D. "Factors Determining the Curriculum in English in Selected Secondary Schools of Texas." Unpublished doctoral dissertation, University of Texas (1953).

3. Morrison, Wilma. "City Schools Start "Operation English" to Lift Standard of Written Language to Higher Level." Oregonian, (Sept. 9, 1951), p. 12.

4. National Council of Teachers of English. The National Interest in the Teaching of English. Champaign, Ill. (1961).

5. Pooley, R. C. and Searles, J. R. The Program of English Instruction in Milwaukee High Schools. Report to the Superintendent of Schools, Milwaukee, Wisconsin, (June, 1956).

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

A STUDY OF INTELLIGENCE AND ACHIEVEMENT AT THE FOURTH-, FIFTH-, AND SIXTH-GRADE LEVELS[a]

HARRY E. ANDERSON, JR.
California Test Bureau
Monterey, California

ALEC J. SLIVINSKE
American Institute for Research
Los Angeles, California

THE PRESENT STUDY is an extension of previous work to investigate the differential relationships regarding school grades, intelligence, and achievement test scores at three grade levels. Anderson (1) predicted reading and language school grades at the fourth-grade level and, in another study (2) language and nonlanguage school grades at the sixth-grade level. More data has been made available on the fourth-grade sample, and a fifth-grade sample has been added so that comparisons can be made among all three grade levels. Moreover, the previous studies used the multiple correlation technique which is well known (e.g., Anderson and Fruchter, 3) as a method for eliminating predictor variables without serious loss of predictive power. The present study uses the method of factor analysis in an effort to explore the fundamental domain of variables at each grade level.

Several authors (e.g., Freeman, 7) have pointed out the need for further data on the California Achievement Test and the California Test of Mental Maturity (the standardized tests used herein) with students below the seventh-grade level, and studies such as the present one, exploring the nature of development at several lower grade levels, have implications for many aspects of education. Barrett (6), for instance, found that underachievement was present as early as the fifth grade, and Shaw and McCuen (11) studied underachievement further with bright children. Scannell (10) was interested in predicting success in college from various grade levels, as low as grade four, and found that the upper grade levels furnished more prediction. Schannell, however, did not investigate the differential characteristics of achievement at the various grade levels. Knief and Stroud (9) found that achievement at the upper elementary grade levels, as indicated by the Iowa Test of Basic Skills, was predicted very well by the Lorge-Thorndike verbal and nonverbal intelligence tests; school grades, however, were not included in the analysis, and the authors emphasized the need for further investigations of verbal and nonverbal intelligence test scores. Gunn (8) was interested in studying single and multiple grade groupings at the fourth-, fifth-, and sixth-grade levels, but encountered major analytic problems. Achievement differences, of course, are extant within grade levels as well as between grade levels, and some studies (e.g., Anderson and Guenter, 4, and Anderson and Whipple, 5), have investigated achievement characteristics within courses. These, and similar studies, may be enhanced by information regarding intelligence-achievement analytic differences between grade levels.

Method

Data for ten variables were collected for fourth-, fifth-, and sixth-grade students in four elementary schools in the Los Angeles area; the sample sizes were, respectively, 112, 165, and 199. The variables, together with appropriate designations used in the study, are as follows: California Test of Mental Maturity (1957 edition), Language (LM), and Nonlanguage (NLM); California Achievement Test (1957 edition), Reading Vocabulary (RV), Reading Comprehension (RC), Arithmetic Reasoning (AR), and Arithmetic Fundamentals (AF); and school grades in Reading (R), Arithmetic (A), Language (L), and Spelling (S). The school grades (scored 15 for an A+, 14 for an A, 13 for an A-, 12 for a B+, etc.) for all samples were those obtained by the students approximately at the same time they took the California Achievement Test so that these standardized achievement scores were not available to the teachers at the time school grades were given. Such is not the case for the

TABLE I

MEANS AND STANDARD DEVIATIONS FOR FOURTH- (N=112), FIFTH- (N=165), AND SIXTH- (N=199) GRADERS ON ALL TEN VARIABLES

Variables	Means			Standard Deviations		
	Grade Four	Grade Five	Grade Six	Grade Four	Grade Five	Grade Six
California Test of Mental Maturity						
Language (LM)	102.5	101.1	99.2	14.3	16.0	16.8
Nonlanguage (NLM)	101.7	100.8	102.2	14.3	20.2	20.1
California Achievement Test						
Reading Vocabulary (RV)	20.8	29.0	34.7	12.5	10.9	10.2
Reading Comprehension (RC)	26.2	39.7	46.6	12.3	14.3	13.3
Arithmetic Reasoning (AR)	17.3	24.6	29.0	9.1	8.1	7.5
Arithmetic Fundamentals (AF)	20.5	30.9	42.6	9.1	9.2	11.5
School Grades						
Reading (R)	10.9	10.7	10.9	2.3	2.0	2.1
Arithmetic (A)	10.7	10.2	10.1	2.4	2.5	2.3
Language (L)	10.7	10.4	10.2	2.0	1.9	2.2
Spelling (S)	10.5	10.2	10.6	2.5	2.6	2.6

TABLE II

VARIABLES' CORRELATIONS FOR FOURTH GRADERS*

	LM	NLM	RV	RC	AR	AF	R	A	L	S
LM		465	289	275	221	088	261	182	353	100
NLM			268	306	213	171	283	259	356	167
RV				815	795	670	492	411	431	478
RC					785	584	545	524	492	529
AR						773	358	363	321	412
AF							209	276	202	302
R								637	630	609
A									601	619
L										555
S										

* Decimals omitted. r = .196 is significant at the .05 level
r = .258 is significant at the .01 level

TABLE III

VARIABLES' CORRELATIONS FOR FIFTH GRADERS*

	LM	NLM	RV	RC	AR	AF	R	A	L	S
LM		573	763	742	721	606	470	440	383	472
NLM			486	606	632	616	230	374	201	268
RV				870	791	503	524	366	436	460
RC					849	638	509	453	439	487
AR						725	356	386	302	401
AF							356	494	378	426
R								657	679	686
A									689	754
L										670
S										

* Decimals omitted. r = .153 is significant at the .05 level
r = .201 is significant at the .01 level

TABLE IV

VARIABLES' CORRELATIONS FOR SIXTH GRADERS*

	LM	NLM	RV	RC	AR	AF	R	A	L	S
LM		523	631	651	608	461	525	314	411	364
NLM			454	556	584	528	364	249	265	308
RV				851	759	582	642	416	563	543
RC					834	695	612	411	561	534
AR						758	539	376	414	480
AF							436	492	354	488
R								559	668	653
A									630	619
L										620
S										

* Decimals omitted. r = .137 is significant at the .05 level
r = .181 is significant at the .01 level

California Test of Mental Maturity scores; subjective information, however, indicates that these scores are only of limited analytic interest to teachers.

Several analyses[b] were conducted with the data from all four schools tobether. This procedure confounds school and teacher effects, but allows for assessment of general underlying relationships. The ten variables were intercorrelated separately for the fourth-grade, fifth-grade, and sixth-grade students. Each of the three correlation matrices was factored by the principal component method, with unity in the diagonals, and an ortho-quartimax rotation was performed on the resulting structure.

Results

Sample means and standard deviations, as shown in Table I, on the standardized variables are consistent with appropriate norms given in the manuals for the 1957 editions of these tests, indicating that our samples are fairly representative of the population on which the norms were constructed. The sample means for the school grades, in all subjects at all grade levels, ranged between B and B-.

The product-moment correlations at the fourth-, fifth-, and sixth-grade levels are shown, respectively, in Tables II, III, and IV. Each of the three correlation matrices was factored by the principal component method with unity throughout the diagonals. Four principal components were extracted in each of the three factor analyses, the factor extraction process being abandoned at this time because of the marked decrement in accountable variance. The principal component matrices are shown in Tables V, VI, and VII, respectively, for the fourth-, fifth-, and sixth-grade analyses. The factors in each of the three analyses were rotated by the ortho-quartimax method, and the rotated factors are shown in Tables VIII, IX, and X.

Fourth-grade factors. Factor I is a CAT Factor with loadings for all four CAT variables ranging from .80 for Reading Comprehension to .93 for Arithmetic Reasoning; the four school grade variables have appreciable loadings on Factor I, also, but markedly lower than the CAT variables. Moreover, the major portion of variance of the school grade variables is located on Factor II which is defined as a School Grades Factor. Reading Vocabulary and Reading Comprehension have Factor II loadings, respectively, of .42 and .29, also, suggesting an appreciable amount of reading variance in the school grades. Factors III and IV are quite specific, respectively, to the CTMM Language and Nonlanguage variables and are so designated. The mental test variables seem quite specific in this analysis, as indicated by the latter two factors, so that the common factor variance can be defined very well in 2-space.

Fifth-grade factors. Factor I is a standardized Test Factor with loadings for five of the standardized test variables ranging from .57 for the CTMM Nonlanguage to .92 for Reading Vocabulary; other noticeable loadings are shown by Reading .35, Arithmetic Fundamentals .34, and Spelling .28, suggesting a latent generality to this factor. The School Grades Factor is again Factor II in this analysis with loadings ranging from .77 for Reading to .88 for Arithmetic; it is noteworthy, however, that Arithmetic Fundamentals, with .73, has most of its common variance on this factor. Factor III is quite specific to NLM. Factor IV is designated as an Arithmetic-Reading bipolar factor with Arithmetic Fundamentals .52, Arithmetic Reasoning .26, and Reading -.30.

A more general factor arises in this analysis, but the school grades remain somewhat distinct. Contrary to results in the fourth-grade analysis, the Arithmetic Fundamentals variable, rather than the CAT reading variables, has the strongest relationship to the School Grades Factor. Factor III is quite specific to the CTMM Nonlanguage variable, but the bipolar fourth factor suggests that language-nonlanguage variances have some degree of inverse relationship; the latter factor is not pronounced, however, to any great degree.

Sixth-grade factors. Factor I is a General Factor with all variables' loadings from .34 (Arithmetic) to .94 (Reading Vocabulary). Factor II is again a School Grades Factor with loadings for all four variables ranging between .55 for Reading to .81 for Arithmetic. Factor III is defined as a bipolar Language-Arithmetic Factor with loadings as follows: CTMM Language .39, Reading .28, Language .26, Arithmetic Fundamentals -.53, and Arithmetic -.21. Factor IV is a Mental Test Factor with the CTMM variables Nonlanguage .78 and Language .32.

The General Factor in the sixth-grade analysis is quite clear, but the school grade variables come out again evenly more distinctly on Factor II. The bipolar language-nonlanguage factor, Factor III, seems stronger here than in the fifth-grade analysis. Finally, the two intelligence subtests emerge together on a separate factor, although their loadings are quite disparate.

Discussion

The factor analyses in the present study suggest something of a difference in the nature of variation in the variables at the three grade levels. The variation in the CAT scores and school grades in the fourth grade seems to be quite general within those variables. Language and nonlanguage types of variation are distinguishable only in the intelligence test scores at this grade level, and these are quite specific. A more general factor is evidenced

TABLE V

PRINCIPAL COMPONENTS FOR FOURTH-GRADE DATA

Variables	Principal Components I	II	III	IV
LM	.408	-.383	-.668	.441
NLM	.447	-.380	-.592	-.538
RV	.854	.321	-.076	.116
RC	.884	.204	.006	.005
AR	.797	.494	-.101	.039
AF	.650	.600	-.103	-.160
R	.734	-.386	.312	.077
A	.713	-.332	.393	-.151
L	.708	-.470	.155	.061
S	.580	-.102	.217	-.011
Sum of Squares	4.82	1.53	1.15	.56

TABLE VI

PRINCIPAL COMPONENTS FOR FIFTH-GRADE DATA

Variables	Principal Components I	II	III	IV
LM	.823	-.248	.142	-.110
NLM	.655	-.423	-.428	-.413
RV	.831	-.241	.430	-.022
RC	.881	-.274	.203	.022
AR	.827	-.429	.044	.176
AF	.758	-.237	-.407	.375
R	.707	.499	.200	-.154
A	.721	.495	-.314	-.055
L	.663	.572	.028	.109
S	.727	.507	-.048	.010
Sum of Squares	5.82	1.69	.74	.39

TABLE VII

PRINCIPAL COMPONENTS FOR SIXTH-GRADE DATA

Variables	Principal Components			
	I	II	III	IV
LM	.722	-.275	.369	-.236
NLM	.625	-.449	-.170	-.550
RV	.860	-.119	.233	.259
RC	.894	-.222	.094	.202
AR	.846	-.345	-.093	.206
AF	.764	-.232	-.481	.168
R	.788	.307	.235	-.075
A	.650	.527	-.337	-.138
L	.714	.494	.190	-.072
S	.731	.427	-.137	.012
Sum of Squares	5.84	1.31	.69	.57

TABLE VIII

ROTATED FACTORS FOR FOURTH-GRADE DATA

Variables	Factors			
	I (CAT)	II (School Grades)	III (LM)	IV (NLM)
LM	.1503	.1504	.9238	.2358
NLM	.1606	.1972	.2307	.9313
RV	.8606	.2896	.1648	.0101
RC	.7964	.4198	.1171	.0509
AR	.9322	.1360	.0608	.0194
AF	.8885	-.0219	-.1325	.1095
R	.2622	.8428	.1119	.0163
A	.2681	.8348	-.1135	.1206
L	.2079	.7978	.2264	.1376
S	.3415	.5263	-.0095	.0214

TABLE IX

ROTATED FACTORS FOR FIFTH-GRADE DATA

Variables	Factors			
	I (Standardized Test)	II (School Grades)	III (NLM)	IV (Arithmetic-Reading)
LM	.8141	.2780	-.1644	-.0652
NLM	.5670	.1463	.7857	.0405
RV	.9193	.2336	-.0982	-.1573
RC	.8994	.2848	.0581	.0151
AR	.8907	.1669	.1170	.2573
AF	.3397	.7313	.1213	.5209
R	.3492	.7700	-.0865	-.3002
A	.1894	.8785	.2306	.0682
L	.2286	.8396	-.1473	-.0128
S	.2820	.8409	-.0042	-.0318

TABLE X

ROTATED FACTORS FOR SIXTH-GRADE DATA

Variables	Factors			
	I (General)	II (School Grades)	III (Language-Arithmetic)	IV (Mental Test)
LM	.7324	.0400	.3895	.3158
NLM	.5587	.0339	-.0386	.7806
RV	.9089	.1350	.1236	-.1266
RC	.9423	.1013	-.0019	.0016
AR	.9176	.0119	-.1943	.0848
AF	.7489	.1613	-.5330	.1628
R	.6236	.5542	.2834	-.0006
A	.3409	.8148	-.2075	.0985
L	.4912	.6924	.2629	-.0714
S	.5085	.6851	-.0798	-.0360

in the fifth-grade analysis, and this is even more salient in the sixth-grade analysis. Moreover, the bipolar language-nonlanguage factor that appeared in the fifth-grade analysis is even stronger in the sixth-grade analysis.

The results of the factor analyses suggest two interesting aspects of intellectual growth and development from the fourth grade to the sixth grade. First, there is greater generality in the variation of intelligence and achievement as the children progress from the fourth to the sixth grade so that the sixth-grade analysis produces a factor that is quite general across all variables. Second, the language and nonlanguage types of variation appear more distinguishable, even displaying something of an inverse relationship, as the students move from the fourth to the sixth grade. These results have implications for other studies in educational achievement. Shaw and McCuen (11), for instance, found that the fourth, fifth, and sixth grades were crucial grades in terms of underachievement and overachievement in bright students. This phenomenon might well be related to the rising importance of numerical-verbal differences reflected in the present study. Longitudinal studies, however, would probably be more revealing in terms of investigating apposite relationships and differential characteristics of academic achievement.

Summary

Data were collected for students in the fourth, fifth, and sixth grades using ten variables as follows: Language and Nonlanguage scores in the California Test of Mental Maturity; Reading Vocabulary, Reading Comprehension, Arithmetic Reasoning, and Arithmetic Fundamentals in the California Achievement Test; school grades in Reading. Arithmetic, Language, and Spelling. The data were factor analyzed and comparisons were made among the factors at the three grade levels.

FOOTNOTES

a. The authors are indebted to several persons for assistance in the data collection: Charles Boulanger, Rose Marie Fong, Mary Gostlin, Grover Holmes, Emily Johnson, Kenneth Johnson, Harry Kirshner, Ruth Libeu, Helen Reuter, Ira Taneman, and Jerry Yerian.

b. The authors express their appreciation to the System Development Corporation for use of their computer facilities.

REFERENCES

1. Anderson, H. E., Jr. "The Prediction of Reading and Language from the California Tests," Educational and Psychological Measurement, XXI (1961), pp. 1035-36.
2. Anderson, H. E., Jr. "A Study of Language and Nonlanguage Achievement," Educational and Psychological Measurement, XXI (1961), pp. 1037-38.
3. Anderson, H. E., Jr., and Fruchter, B. "Some Multiple Correlation and Predictor Selection Methods," Psychometrika, XXV (1960), pp. 59-76.
4. Anderson, H. E., Jr., and Guenter, F. "Aptitude and Course Achievement in a Short Missile Technician Course," Journal of Experimental Education, XXIV (1961), pp. 291-8.
5. Anderson, H. E., Jr., and Whipple, J. "Course Achievement of Students with Unsatisfactory Academic Averages in Basic Electronics," Staff Memorandum. Fort Bliss: USAADHRU (HumRRO), 1958.
6. Barrett, H.O. "An Intensive Study of Thirty-Two Gifted Children," Personnel Guidance Journal, XXXVI (1957), pp. 192-4.
7. Freeman, F.S. "California Test of Mental Maturity, 1957 Edition," in O. K. Buros, The Fifth Mental Measurements Yearbook (New Jersey: The Gryphon Press, 1959).
8. Gunn, E. M. An Analysis of the CAT Scores of Regular Grade and Multi-grade Students in the Saugus Elementary School District, unpublished M. A. thesis, University of Southern California, 1959.
9. Knief, L. M., and Stroud, J. B. "Intercorrelations Among Various Intelligence, Achievement, and Social Class Scores," Journal of Educational Psychology, L (1959), pp. 117-20.
10. Scannell, D. P. "Prediction of College Success from Elementary and Secondary School Performance," Journal of Educational Psychology, LI (1960), pp. 130-34.
11. Shaw, M. C., and McCuen, J. T. "The Onset of Academic Underachievement in Bright Children," Journal of Educational Psychology, LI (1960), pp. 103-08.

THE JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 31, Number 4, Summer 1963)

EMPIRICAL RELATIONSHIPS AMONG MODES OF TESTING, MODES OF INSTRUCTION AND READING LEVELS: IN SIXTH-GRADE SOCIAL STUDIES [1,2]

WILLIAM L. BURR
Harcourt, Brace and World, Inc.
Tarrytown, New York

THE CENTRAL PURPOSES of this study were to determine relationships among sets of test scores resulting from three modes of item presentation and to determine the influence of certain supposed intervening variables on such relationships. The modes of item presentation were: traditional paper and pencil; paced, projected silent; and paced, projected sound. The hypothesized intervening variables were mode of instruction received by subjects (televised and traditional), and general academic aptitude.

Problem

It was hypothesized that there would be no significant difference at the .05 level of significance between the mean scores of the traditionally instructed group and the television-instructed group when test items were presented (a) in the usual booklet form, (b) when items were presented one-at-a-time by projecting them on a screen, and (c) when items were presented as in (b) above except that each item was read aloud when it appeared on the screen. Additional null hypotheses dealt with the interactions of the above variables and the rank order of subjects with differing levels of academic aptitude.

Method

Subjects. All Ss were white, sixth-grade public school students. All Ss in the experimental (E) group came from three sixth-grade social studies classes in one school. The Ss in the control (C) group came from three sixth-grade classes in a neighboring school. Both schools functioned under the same Board of Public Instruction, were comparable with regard to socio-economic class and ability of students, educational facilities, and teacher effectiveness so far as it could be determined. Ss in the E group had received televised instruction in social studies for fifteen months prior to the experiment. In addition, these Ss received a televised enrichment hour and a reading hour each week. Their other school work was given under typical conditions to class groups of the usual size. Ss in the C group received social studies instruction in a comparable amount under normal teaching conditions and none of the students in the school from which they were drawn received televised instruction. In each school, students were assigned to class groups on a random basis. The reading ability and mental ability ranges were essentially the same for both the E and C groups. The E and C groups were formed by matching pairs of students on intelligence, composite prognostic score (a score which consisted of weighted achievement and intelligence scores that was found to be predictive of future academic success), and sex. Fifty-three pairs of matched Ss were formed. Students for whom matches could not be found were discarded prior to the analysis of the data.

Tests. The experimental tests were the Social Studies subtests of the Metropolitan Achievement Test Series. [3] The Social Studies subtests consisted of two parts, one dealing with social studies information and the other dealing with study skills. The information part was highly verbal; the study skills part required both verbal and graphic skills. The normal administration of these tests consisted of administering them in booklet form according to

TABLE I

ANALYTIC DESIGN

General Academic Aptitude	Test Presentations	Groups C	Groups E
Low	Normal Projected Silent Projected Sound	n = 20	n = 20
Average	Normal Projected Silent Projected Sound	n = 15	n = 13
High	Normal Projected Silent Projected Sound	n = 18	n = 20

TABLE II

ANALYSIS OF VARIANCE OF VERBAL ITEMS

Source of Variation	df	MS	F
Instructional Groups (G)	1	11.10	.12
Academic Aptitude (A)	2	8486.08	90.96*
G X A	2	85.58	.92
Error	100	93.29	
Test Presentations (T)	2	.005	.00
G X T	2	2.99	.14
A X T	4	73.30	3.48*
G X A X T	4	121.78	5.77*
Error	200	21.09	
Total	317		

* Significant at .05

TABLE III

ANALYSIS OF VARIANCE OF GRAPHIC ITEMS

Source of Variation	df	MS	F
Instructional Groups (G)	1	14.54	.13
Academic Aptitude (A)	2	6678.15	60.43*
G X A	2	287.51	2.60
Error	100	110.51	
Test Presentations (T)	2	.23	.00
G X T	2	16.36	.55
A X T	4	11.57	.39
G X A X T	4	54.29	1.84
Error	200	29.50	
Total	317		

* Significant at .05

the directions contained in the test manual. For the projected administrations, the directions, sample items, and test items were reproduced on 35 mm slides, one item to a slide. For the projected, sound administration a tape recorder and slide projector were electronically so synchronized that when a test, item, directions, etc., appeared on the screen, the reading of such material was broadcast immediately by means of the precut tape.

Testing Procedure. Each class of students was administered the tests in their classroom. There were eighteen testing sessions in all. Each class was tested on three consecutive days. Each class responded to three forms of the Social Studies subtests of the Metropolitan Achievement Test Series, each administered under different conditions. The order of conditions was randomized over all classes.

The control administration was given according to publishers' directions. The projected silent test was administered with these time limits: verbal items, 15 seconds per item; and graphic items, 50 seconds per item. The projected sound test was administered with the same times as the projected silent test, the difference being the audio supplement.

The experimental testing times were established in the following way. A pilot study was conducted with one representative class of sixth-grade students to determine the exposure time interval to use in pacing the students through the projected tests. After an analysis of the results, the decision was made to use the exposure time intervals indicated in the preceding paragraph.

The reader should note that less time for response was given to the students on the projected tests than on the normal tests. Consequently, the time variable could conceivably be a factor in the performance of the students on the tests.

Analysis

The analytic design for this study is depicted in Table I.

The results for verbal and graphic items were separately subjected to an analysis of variance shown in Table I and Table II.

The Newman-Keuls sequential range test was used to make multiple, non-significance tests on the differences between the ranked means of the significant interactions.

A separate examination was made of the distribution of signed differences between projected silent and projected sound test scores and reading ability.

Conclusions

Only the general academic aptitude source of variation, of the main effects, is significant at the .05 level. The verbal item analysis indicates that two interactions are significant at the .05 level.

The only significant trend discovered by the use of the Newman-Keuls range test was a result of general academic aptitude grouping. This finding was consistent with the results shown in Table II.

The separate examination of the distribution of signed differences between projected silent and projected sound test scores and reading ability, suggest that if audio-reinforcement of visually projected test items does affect test performance, then such effects will be found below a reading grade equivalent of 4.7 and above 8.7.

It is concluded that either silent or audio-reinforced projected achievement testing will produce results comparable to paper and pencil testing for either traditionally instructed or television-instructed groups and will, in addition, utilize less testing time.

There were no significant differences in performance of the two instructional groups when tested on two types of items, verbal and graphic.

Subsequent studies might investigate the possibility of the development of viewing skills through the home viewing of television, which was not investigated in this study.

FOOTNOTES

1. Taken from a doctoral dissertation submitted to the School of Education, Florida State University, in partial fulfillment of the requirements for a Doctor of Education degree.
2. This study was supported in part by a grant from the United States Department of Health, Education and Welfare, Office of Education Title VII, National Defense Education Act of 1958, Grant Number 7-08-075.00 and the Florida State University.
3. The author is indebted to Harcourt, Brace and World, Inc., for permission to use experimentally the Social Studies subtests of Forms A, B and C of the Metropolitan Achievement Test Series.